

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number  
**WO 01/79231 A2**

- (51) International Patent Classification<sup>7</sup>: **C07H**
- (21) International Application Number: PCT/US01/12300
- (22) International Filing Date: 16 April 2001 (16.04.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/197,330 14 April 2000 (14.04.2000) US
- (71) Applicant (*for all designated States except US*): **GENAIS-SANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **BENTIVEGNA, Steven, C.** [US/US]; 49 Greenview Terrace, Middletown, CT 06457 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KLIEM, Stephanie, E.** [DE/US]; 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **NANDABALAN, Krishnan** [IN/US]; 228 Village Pond Road, Guilford, CT 06437 (US).
- (74) Agents: **SHTIVELBAND, Inna** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: HAPLOTYPES OF THE NPR1 GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human natriuretic peptide receptor A/guanylate cyclase A (atrio-natriuretic peptide receptor A) (NPR1) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the NPR1 gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the NPR1 gene in an individual are also disclosed. Polynucleotides containing one or more of the NPR1 polymorphisms disclosed herein are also described.



WO 01/79231 A2

## HAPLOTYPES OF THE NPR1 GENE

## RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/197,330 filed April 14, 2000.

## FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

## BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the

population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyen PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of hypertension is the natriuretic peptide receptor A/guanylate cyclase A (atriuretic peptide receptor A) (NPR1) gene or its encoded product. NPR1, also known as NPRA, is a receptor that binds to atrial natriuretic peptides (ANP). ANP produced in the heart causes vasodilation and natriuresis, which are important for the

regulation of blood pressure. Mice lacking functional NPR1 have elevated blood pressures and hearts exhibiting marked hypertrophy with interstitial fibrosis resembling that seen in human hypertensive heart disease (Oliver et al., *Proc. Natl. Acad. Sci. U. S. A* 1997; 94:14730-14735). The binding of ANP to the extracellular domain of NPR1 activates the receptor guanylate cyclase to synthesize cGMP. (Lowe, *Biochemistry* 1992; 31:10421-10425). Increases in the levels of cGMP causes the downregulation of NPR1 mRNA, thus allowing NPR1 to autoregulate its own transcription (Cao et al., *Am. J. Physiol* 1998; 275:F119-F125).

Serum testosterone levels tend to be lower in hypertensive males than in normal males. Pandey et al. (*Endocrinology* 1999; 140:5112-5119) studied the influence of NPR1 on serum testosterone levels in male hypertensive rats lacking a functional NPR1 gene, wild type animals expressing two copies, and those expressing four copies of the NPR1 gene. The animals with four copies of NPR1 gene had higher levels of testosterone than those with two copies of the gene. The NPR1 knockout mice had testosterone levels lower than the two-copy mice. Also, Leydig cells lacking NPR1, did not show ANP-stimulated cGMP accumulation and had no ANP-dependent testosterone production. This study establishes the role of NPR1 in testicular steroidogenesis, and shows a relationship between hypertension associated with decreased NPR1 and low testosterone levels.

ANP has also been shown to inhibit the agonist stimulated activity of mitogen-activated protein kinase/extracellular signal regulated kinase 2 (MAPK/ERK2). This inhibitory effect of ANP was reversed on treatment with NPR1 antagonist, suggesting that the ANP/NPR1 system negatively regulates MAPK/Erk2 (Pandey et al., *Biochem. Biophys. Res. Commun.* 2000; 271:374-379).

The natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) gene is located on chromosome 1q21-q22 and contains 22 exons that encode a 1061 amino acid protein. Reference sequences for the NPR1 gene (Genaissance Reference No. 1568505; SEQ ID NO:1), coding sequence (GenBank Accession No:NM\_000906.1), and protein are shown in Figures 1, 2 and 3, respectively.

Because of the potential for variation in the NPR1 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the NPR1 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of NPR1 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

## SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 21 novel polymorphic sites in the NPR1 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 730 (PS1), 731 (PS2), 811 (PS3), 822 (PS4), 1235 (PS5), 1351 (PS6), 2184 (PS7), 2472 (PS8), 2979 (PS9), 4345 (PS10), 5290 (PS11), 5537 (PS12), 6900 (PS13), 7410 (PS14), 7947 (PS15), 9313

(PS16), 9619 (PS17), 9675 (PS18), 9904 (PS19), 10004 (PS20) and 11062 (PS21). The polymorphisms at these sites are guanine or adenine at PS1, guanine or cytosine at PS2, cytosine or thymine at PS3, cytosine or adenine at PS4, guanine or cytosine at PS5, cytosine or thymine at PS6, thymine or cytosine at PS7, adenine or guanine at PS8, guanine or cytosine at PS9, thymine or adenine at PS10, thymine or cytosine at PS11, guanine or adenine at PS12, guanine or adenine at PS13, adenine or thymine at PS14, cytosine or thymine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, adenine or thymine at PS18, cytosine or thymine at PS19, guanine or adenine at PS20 and cytosine or thymine at PS21. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-21 in the NPR1 gene, which are shown below in Tables 5 and 4, respectively. Each of these NPR1 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the NPR1 gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the NPR1 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21 in both copies of the NPR1 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel NPR1 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel NPR1 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the NPR1 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the NPR1 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's NPR1 gene is defined by one of the NPR1 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's NPR1 gene are defined by one of the NPR1 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the NPR1 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with NPR1 activity, e.g., hypertension.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate NPR1 as a candidate target for treating a specific condition or disease predicted to be associated with NPR1 activity. Determining for a particular population the frequency of one or more of the individual NPR1 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue NPR1 as a target for treating the specific disease of interest. In particular, if variable NPR1 activity is associated with the disease, then one or more NPR1 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed NPR1 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable NPR1 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any NPR1 haplotype or haplotype pair, apply the information derived from detecting NPR1 haplotypes in an individual to decide whether modulating NPR1 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting NPR1 to treat a specific condition or disease predicted to be associated with NPR1 activity. For example, detecting which of the NPR1 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent NPR1 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular NPR1 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the NPR1 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with NPR1 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the NPR1 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute NPR1 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a NPR1 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any NPR1 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a NPR1 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the NPR1 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the

NPR1 genotype or haplotype in a reference population. A higher frequency of the NPR1 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the NPR1 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the NPR1 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for hypertension.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the NPR1 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, cytosine at PS5, thymine at PS6, cytosine at PS7, guanine at PS8, cytosine at PS9, adenine at PS10, cytosine at PS11, adenine at PS12, adenine at PS13, thymine at PS14, thymine at PS15, adenine at PS16, adenine at PS17, thymine at PS18, thymine at PS19, adenine at PS20 and thymine at PS21.

A particularly preferred polymorphic variant is an isogene of the NPR1 gene. A NPR1 isogene of the invention comprises guanine or adenine at PS1, guanine or cytosine at PS2, cytosine or thymine at PS3, cytosine or adenine at PS4, guanine or cytosine at PS5, cytosine or thymine at PS6, thymine or cytosine at PS7, adenine or guanine at PS8, guanine or cytosine at PS9, thymine or adenine at PS10, thymine or cytosine at PS11, guanine or adenine at PS12, guanine or adenine at PS13, adenine or thymine at PS14, cytosine or thymine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, adenine or thymine at PS18, cytosine or thymine at PS19, guanine or adenine at PS20 and cytosine or thymine at PS21. The invention also provides a collection of NPR1 isogenes, referred to herein as a NPR1 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a NPR1 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 5, adenine at a position corresponding to nucleotide 16, cytosine at a position corresponding to nucleotide 429, thymine at a position corresponding to nucleotide 545, cytosine at a position corresponding to nucleotide 1023 and thymine at a position corresponding to nucleotide 2406. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a NPR1 isogene defined by haplotypes 1-6 and 8-14.

Polynucleotides complementary to these NPR1 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the NPR1-gene will be useful in studying the expression and function of NPR1, and in expressing NPR1 protein for use in screening for candidate drugs to treat diseases related to NPR1 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express NPR1 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the NPR1 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of leucine at a position corresponding to amino acid position 2, serine at a position corresponding to amino acid position 6, valine at a position corresponding to amino acid position 182 and isoleucine at a position corresponding to amino acid position 341. A polymorphic variant of NPR1 is useful in studying the effect of the variation on the biological activity of NPR1 as well as on the binding affinity of candidate drugs targeting NPR1 for the treatment of hypertension.

The present invention also provides antibodies that recognize and bind to the above polymorphic NPR1 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the NPR1 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the NPR1 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against NPR1 protein, and for testing the efficacy of therapeutic agents and compounds for hypertension in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the NPR1 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the NPR1 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing NPR1 haplotypes organized according to their evolutionary relationships.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the NPR1 gene (Genaissance Reference No. 1568505; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([ or ]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:109 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and

W= A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the NPR1 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the NPR1 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the NPR1 gene. As described in more detail below, the inventors herein discovered 15 isogenes of the NPR1 gene by characterizing the NPR1 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (CA) (22 individuals), African descent (AF) (20 individuals), Asian (AS) (20 individuals), or Hispanic/Latino (HL) (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The NPR1 isogenes present in the human reference population are defined by haplotypes for 21 polymorphic sites in the NPR1 gene, all of which are believed to be novel. The NPR1 polymorphic sites identified by the inventors are referred to as PS1-21 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21. Using the genotypes identified in the Index Repository for PS1-21 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the NPR1 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the NPR1 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether NPR1 is a suitable target for drugs to treat hypertension, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

**Allele** - A particular form of a genetic locus, distinguished from other forms by its particular

nucleotide sequence.

**Candidate Gene** – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

**Gene** - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

**Genotype** – An unphased 5' to 3' sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

**Full-genotype** – The unphased 5' to 3' sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

**Sub-genotype** – The unphased 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

**Genotyping** – A process for determining a genotype of an individual.

**Haplotype** – A 5' to 3' sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

**Full-haplotype** – The 5' to 3' sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

**Sub-haplotype** – The 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

**Haplotype pair** – The two haplotypes found for a locus in a single individual.

**Haplotyping** – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

**Haplotype data** - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

**Isoform** – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

**Isogene** – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

**Isolated** – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with

the methods of the present invention.

**Locus** - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

**Naturally-occurring** – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

**Nucleotide pair** – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

**Phased** – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

**Polymorphic site (PS)** – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

**Polymorphic variant** – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

**Polymorphism** – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

**Polymorphism data** – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

**Polymorphism Database** – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

**Polynucleotide** – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

**Population Group** – A group of individuals sharing a common ethnogeographic origin.

**Reference Population** – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

**Single Nucleotide Polymorphism (SNP)** – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

**Subject** – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

**Treatment** - A stimulus administered internally or externally to a subject.

**Unphased** - As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the NPR1 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel NPR1 polymorphisms and haplotypes identified herein.

The compositions comprise at least one NPR1 genotyping oligonucleotide. In one embodiment, a NPR1 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a NPR1 polynucleotide, i.e., a NPR1 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-NPR1 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the NPR1 gene using the polymorphism information provided herein in conjunction with the known sequence information for the NPR1 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect"

or “complete” complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is “substantially complementary” to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in *Molecular Cloning, A Laboratory Manual*, 2<sup>nd</sup> Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5' end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., “Genetic Prediction of Hemophilia A” in *PCR Protocols, A Guide to Methods and Applications*, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7<sup>th</sup> or 8<sup>th</sup> position in a 15mer, the 8<sup>th</sup> or 9<sup>th</sup> position in a 16mer, and the 10<sup>th</sup> or 11<sup>th</sup> position in a 20mer). An ASO primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting NPR1 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

GATGCCTRGGACCGG (SEQ ID NO:4) and its complement,  
 ATGCCTGSGACCGGC (SEQ ID NO:5) and its complement,  
 GCCATGCGGGGCCC (SEQ ID NO:6) and its complement,  
 GCCCCGGMGCCCCGC (SEQ ID NO:7) and its complement,  
 CCCCCGCSCTGGGCT (SEQ ID NO:8) and its complement,  
 CGCCAAGYGCTCATG (SEQ ID NO:9) and its complement,  
 CTGACTCYCCGTCTT (SEQ ID NO:10) and its complement,  
 CCTCAGCRTCTGAAA (SEQ ID NO:11) and its complement,  
 TCACCATSGAGGATG (SEQ ID NO:12) and its complement,  
 CTCCTACWTCCCCC (SEQ ID NO:13) and its complement,  
 GCTCCTGYCCCATGC (SEQ ID NO:14) and its complement,  
 GTGATGTRGGGGGTT (SEQ ID NO:15) and its complement,  
 ACTTGCTRTGTGACC (SEQ ID NO:16) and its complement,  
 ATAAGGCWGGATAAG (SEQ ID NO:17) and its complement,  
 TCGGGGAYGCAAGGG (SEQ ID NO:18) and its complement,  
 GACTACCRACCTCTG (SEQ ID NO:19) and its complement,  
 CATTGCTRCCAGTGA (SEQ ID NO:20) and its complement,  
 CCATCTCWGCTGGTT (SEQ ID NO:21) and its complement,  
 CGTTGCGYAAATTTA (SEQ ID NO:22) and its complement,  
 CAGCAGTRGCAGAGG (SEQ ID NO:23) and its complement, and  
 CACCAGAYCTGCCTT (SEQ ID NO:24) and its complement.

A preferred ASO primer for detecting NPR1 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

GCGCCTGATGCCTRG (SEQ ID NO:25); CAGCGGCCGGTCCYA (SEQ ID NO:26);  
 CGCCTGATGCCTGSG (SEQ ID NO:27); TCAGCGGCCGGTCSC (SEQ ID NO:28);  
 GCTGAGGCCATGCGY (SEQ ID NO:29); GCGCCGGGGCCCCRG (SEQ ID NO:30);  
 GCCGGGGCCCCGGMG (SEQ ID NO:31); GAGCCAGCGGGGCKC (SEQ ID NO:32);  
 CCGGCGCCCCGGCSC (SEQ ID NO:33); CACCGAAGCCCAGSG (SEQ ID NO:34);  
 TGGGAGCGCCAAGYG (SEQ ID NO:35); GTAGAGCATGAGCRC (SEQ ID NO:36);  
 CTCTCTCTGACTCYC (SEQ ID NO:37); TGGAGAAAGACGGRG (SEQ ID NO:38);  
 TGTGTCCCTCAGCRT (SEQ ID NO:39); GAATTCTTTTTCAGAYG (SEQ ID NO:40);  
 TCAACTTCACCATSG (SEQ ID NO:41); CCAGGCCATCCTCSA (SEQ ID NO:42);  
 CTCACCCTCCTACWT (SEQ ID NO:43); GCTGTGGGGGGGAWG (SEQ ID NO:44);  
 GTAGGTGCTCCTGYC (SEQ ID NO:45); CCCTCAGCATGGGRC (SEQ ID NO:46);  
 TGAGCTGTGATGTRG (SEQ ID NO:47); TCACTCAACCCCCYA (SEQ ID NO:48);  
 GCTTTCACTTGCTRT (SEQ ID NO:49); GCTCAAGGTCACAYA (SEQ ID NO:50);  
 ACAAAGATAAGGCWG (SEQ ID NO:51); CCCTGCCTTATCCWG (SEQ ID NO:52);  
 GGGCCCTCGGGGAYG (SEQ ID NO:53); CAGTCTCCCTTGCRG (SEQ ID NO:54);  
 CTCGGTGACTACCRA (SEQ ID NO:55); GTGGGTCAGAGGTYG (SEQ ID NO:56);  
 TTGACCCATTGCTRC (SEQ ID NO:57); GACTGGTCACTGGYA (SEQ ID NO:58);  
 CCCCTGCCATCTCWG (SEQ ID NO:59); TGGGGCAACCAGCWG (SEQ ID NO:60);  
 GCCTGACGTTGCGYA (SEQ ID NO:61); ACCTGTAAATTTTC (SEQ ID NO:62);  
 TTATCCCAGCAGTRG (SEQ ID NO:63); GGTCTCCCTCTGCGYA (SEQ ID NO:64);  
 GGATCCCACCAGAYC (SEQ ID NO:65); and AACCAGAAGGCAGRT (SEQ ID NO:66).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the

novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting NPR1 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

CCTGATGCCT (SEQ ID NO: 67); CGGCCGGTCC (SEQ ID NO: 68);  
 CTGATGCCTG (SEQ ID NO: 69); GCGGCCGGTC (SEQ ID NO: 70);  
 GAGGCCATGC (SEQ ID NO: 71); CCGGGGCCCC (SEQ ID NO: 72);  
 GGGGCCCCGG (SEQ ID NO: 73); CCAGCGGGGC (SEQ ID NO: 74);  
 GCGCCCCGGC (SEQ ID NO: 75); CGAAGCCCAG (SEQ ID NO: 76);  
 GAGCGCCAAG (SEQ ID NO: 77); GAGCATGAGC (SEQ ID NO: 78);  
 TCTCTGACTC (SEQ ID NO: 79); AGAAAGACGG (SEQ ID NO: 80);  
 GTCCCTCAGC (SEQ ID NO: 81); TTCTTTCAGA (SEQ ID NO: 82);  
 ACTTCACCAT (SEQ ID NO: 83); GGCCATCCTC (SEQ ID NO: 84);  
 ACCCTCCTAC (SEQ ID NO: 85); GTGGGGGGGA (SEQ ID NO: 86);  
 GGTGCTCCTG (SEQ ID NO: 87); TCAGCATGGG (SEQ ID NO: 88);  
 GCTGTGATGT (SEQ ID NO: 89); CTCAACCCCC (SEQ ID NO: 90);  
 TTCATTGCT (SEQ ID NO: 91); CAAGGTCACA (SEQ ID NO: 92);  
 AAGATAAGGC (SEQ ID NO: 93); TGCCTTATCC (SEQ ID NO: 94);  
 CCCTCGGGGA (SEQ ID NO: 95); TCTCCCTTGC (SEQ ID NO: 96);  
 GGTGACTACC (SEQ ID NO: 97); GGTCAGAGGT (SEQ ID NO: 98);  
 ACCCATGCT (SEQ ID NO: 99); TGGTCACTGG (SEQ ID NO: 100);  
 CTGCCATCTC (SEQ ID NO: 101); GGCAACCAGC (SEQ ID NO: 102);  
 TGACGTTGCG (SEQ ID NO: 103); TGTAAATTT (SEQ ID NO: 104);  
 TCCCAGCAGT (SEQ ID NO: 105); CTCCCTCTGC (SEQ ID NO: 106);  
 TCCCACCAGA (SEQ ID NO: 107); and CAGAAGGCAG (SEQ ID NO: 108).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

NPR1 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized NPR1 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as

hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the NPR1 gene in an individual. As used herein, the terms "NPR1 genotype" and "NPR1 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the NPR1 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the NPR1 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21 in the two copies to assign a NPR1 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-21.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the NPR1 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5' and 3' untranslated regions. If a NPR1 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the NPR1 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21 in that copy to assign a NPR1 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the NPR1 gene or fragment such as one of the methods described above for preparing NPR1 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only

provide haplotype information on one of the two NPR1 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional NPR1 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the NPR1 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-21 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the NPR1 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's NPR1 gene, the phased sequence of nucleotides present at each of PS1-21. The present invention also contemplates that typically only a subset of PS1-21 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a NPR1 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21 in each copy of the NPR1 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-21 in each copy of the NPR1 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the NPR1 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in

individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the NPR1 gene of an individual may also be determined by

hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, *P. Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in *Molecular Diagnosis of Genetic Diseases*, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruano et al., *Nucl. Acids Res.* 17:8392, 1989; Ruano et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's NPR1 haplotype pair is predicted from its NPR1 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a NPR1 genotype for the individual at two or more NPR1 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing NPR1 haplotype

pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the NPR1 haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a  $q\%$  chance of not missing a haplotype that exists in the population at a  $p\%$  frequency of occurring in the reference population, the number of individuals ( $n$ ) who must be sampled is given by  $2n = \log(1-q)/\log(1-p)$  where  $p$  and  $q$  are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., *Principles of Population Genomics*, Sinauer Associates (Sunderland, MA), 3<sup>rd</sup> Ed., 1997) postulates that the frequency of finding the haplotype pair  $H_1 / H_2$  is equal to  $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$  if  $H_1 \neq H_2$  and  $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$  if  $H_1 = H_2$ . A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a NPR1 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is

consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaisance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a NPR1 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel NPR1 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for NPR1 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a NPR1 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular NPR1 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that NPR1 genotype, haplotype or haplotype pair. Preferably, the NPR1 genotype, haplotype, or haplotype pair being compared in the

trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting NPR1 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a NPR1 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the NPR1 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and NPR1 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their NPR1 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the NPR1 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between NPR1 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2<sup>nd</sup> Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the NPR1 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of NPR1 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the NPR1 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The

diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the NPR1 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying NPR1 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the NPR1 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant NPR1 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21. Similarly, the nucleotide sequence of a variant fragment of the NPR1 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the NPR1 gene, which is defined by haplotype 7, (or other reported NPR1 sequences) or to portions of the reference sequence (or other reported NPR1 sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of adenine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, cytosine at PS5, thymine at PS6, cytosine at PS7, guanine at PS8, cytosine at PS9, adenine at PS10, cytosine at PS11, adenine at PS12, adenine at PS13, thymine at PS14, thymine at PS15, adenine at PS16, adenine at PS17, thymine at PS18, thymine at PS19, adenine at PS20 and thymine at PS21. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the NPR1 gene which is defined by any one of haplotypes 1-6 and 8-14 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the NPR1 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

NPR1 isogenes may be isolated using any method that allows separation of the two "copies" of the NPR1 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific

oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides NPR1 genome anthologies, which are collections of NPR1 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A NPR1 genome anthology may comprise individual NPR1 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the NPR1 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred NPR1 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded NPR1 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant NPR1 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as *E. coli*, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably

mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the NPR1 gene will produce NPR1 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a NPR1 cDNA comprising a nucleotide sequence which is a polymorphic variant of the NPR1 reference coding sequence shown in Figure 2. Thus, the invention also provides NPR1 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 5, adenine at a position corresponding to nucleotide 16, cytosine at a position corresponding to nucleotide 429, thymine at a position corresponding to nucleotide 545, cytosine at a position corresponding to nucleotide 1023 and thymine at a position corresponding to nucleotide 2406. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a NPR1 isogene defined by haplotypes 1-6 and 8-14. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized NPR1 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a NPR1 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the NPR1 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the NPR1 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the NPR1 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for

therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular NPR1 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the NPR1 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular NPR1 isogene. Expression of a NPR1 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3' untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions -10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of NPR1 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of NPR1 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2' O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference NPR1 amino acid sequence shown in Figure 3. The location of a variant amino acid in a NPR1 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A NPR1 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of leucine at a position corresponding to amino acid position 2, serine at a position corresponding to amino acid position 6, valine at a position corresponding to amino acid position 182 and isoleucine at a position corresponding to amino acid position 341. The invention specifically excludes amino acid sequences identical to those previously identified for NPR1, including SEQ ID NO:3, and previously described fragments thereof. NPR1 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a NPR1 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table

5.

Table 2. Novel Polymorphic Variants of NPR1  
Polymorphic Amino Acid Position and Identities  
Variant

Number	2	6	182	341
1	P	R	A	I
2	P	R	V	M
3	P	R	V	I
4	P	S	A	M
5	P	S	A	I
6	P	S	V	M
7	P	S	V	I
8	L	R	A	M
9	L	R	A	I
10	L	R	V	M
11	L	R	V	I
12	L	S	A	M
13	L	S	A	I
14	L	S	V	M
15	L	S	V	I

The invention also includes NPR1 peptide variants, which are any fragments of a NPR1 protein variant that contain one or more of the amino acid variations shown in Table 2. A NPR1 peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such NPR1 peptide variants may be useful as antigens to generate antibodies specific for one of the above NPR1 isoforms. In addition, the NPR1 peptide variants may be useful in drug screening assays.

A NPR1 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant NPR1 genomic and cDNA sequences as described above. Alternatively, the NPR1 protein variant may be isolated from a biological sample of an individual having a NPR1 isogene which encodes the variant protein. Where the sample contains two different NPR1 isoforms (i.e., the individual has different NPR1 isogenes), a particular NPR1 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular NPR1 isoform but does not bind to the other NPR1 isoform.

The expressed or isolated NPR1 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the NPR1 protein as discussed further below. NPR1 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be

used.

A polymorphic variant NPR1 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric NPR1 protein. The non-NPR1 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the NPR1 and non-NPR1 portions so that the NPR1 protein may be cleaved and purified away from the non-NPR1 portion.

An additional embodiment of the invention relates to using a novel NPR1 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known NPR1 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The NPR1 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a NPR1 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the NPR1 protein(s) of interest and then washed. Bound NPR1 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel NPR1 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the NPR1 protein.

In yet another embodiment, when a particular NPR1 haplotype or group of NPR1 haplotypes encodes a NPR1 protein variant with an amino acid sequence distinct from that of NPR1 protein isoforms encoded by other NPR1 haplotypes, then detection of that particular NPR1 haplotype or group of NPR1 haplotypes may be accomplished by detecting expression of the encoded NPR1 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel NPR1 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The NPR1 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the NPR1 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the NPR1 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition

which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the NPR1 protein variant from solution as well as react with NPR1 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect NPR1 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel NPR1 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the NPR1 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., *Principles and Practice of Immunoassay*, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; *Current Protocols in Molecular Biology*, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in *Methods in Immunodiagnosis*, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, *Methods in Immunology*, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., *Principles and Practice of Immunoassay*, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, *J. Clin. Chem. Clin. Biochem.*, 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in *Current Protocols in Molecular Biology*, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, *Nature*, 256:495-497; Campbell *Monoclonal Antibody Technology*, the *Production and Characterization of Rodent and Human Hybridomas*, 1985, In: *Laboratory Techniques in Biochemistry and Molecular Biology*, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, *Science*, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 *Proc. Natl. Acad. Sci. USA* 86:10029).

Effect(s) of the polymorphisms identified herein on expression of NPR1 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the NPR1 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into NPR1 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired NPR1 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the NPR1 isogene is introduced into a cell in such a way that it recombines with the endogenous NPR1 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired NPR1 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the NPR1 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the NPR1 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant NPR1 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the NPR1 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human NPR1 isogene and producing human NPR1 protein can be used as biological models for studying diseases related to

abnormal NPR1 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel NPR1 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel NPR1 isogenes; an antisense oligonucleotide directed against one of the novel NPR1 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel NPR1 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel NPR1 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the NPR1 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations).

The NPR1 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

### EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

#### EXAMPLE 1

This example illustrates examination of various regions of the NPR1 gene for polymorphic sites.

##### Amplification of Target Regions

The following target regions of the NPR1 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	263-285	complement of 958-937	696 nt
Fragment 2	538-558	complement of 1261-1240	724 nt
Fragment 3	911-930	complement of 1498-1478	588 nt
Fragment 4	911-930-	complement of 1744-1722	834 nt
Fragment 5	2026-2047	complement of 2681-2660	656 nt
Fragment 6	2681-2702	complement of 3178-3155	498 nt
Fragment 7	4022-4043	complement of 4471-4450	450 nt
Fragment 8	4857-4882	complement of 5438-5415	582 nt
Fragment 9	5254-5275	complement of 5674-5651	421 nt
Fragment 10	6450-6473	complement of 6992-6970	543 nt
Fragment 11	7304-7327	complement of 7809-7787	506 nt
Fragment 12	7650-7672	complement of 8071-8047	422 nt
Fragment 13	8529-8550	complement of 8960-8939	432 nt
Fragment 14	9148-9170	complement of 9701-9679	554 nt
Fragment 15	9565-9588	complement of 10193-10171	629 nt
Fragment 16	10349-10371	complement of 11094-11072	746 nt
Fragment 17	10734-10755	complement of 11190-11166	457 nt
Fragment 18	10954-10976	complement of 11406-11384	453 nt
Fragment 19	11329-11351	complement of 11917-11895	589 nt
Fragment 20	11668-11689	complement of 12281-12303	636 nt
Fragment 21	14605-14627	complement of 15074-15053	470 nt
Fragment 22	14858-14880	complement of 15340-15320	483 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 $\mu$ l
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 $\mu$ l
100 ng of human genomic DNA	= 1 $\mu$ l
10 mM dNTP	= 0.4 $\mu$ l
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 $\mu$ l
Forward Primer (10 $\mu$ M)	= 0.4 $\mu$ l
Reverse Primer (10 $\mu$ M)	= 0.4 $\mu$ l
Water	= 6.6 $\mu$ l

## Amplification profile:

97°C - 2 min.	1 cycle
97°C - 15 sec.	} 10 cycles
70°C - 45 sec.	
72°C - 45 sec.	
97°C - 15 sec.	} 35 cycles
64°C - 45 sec.	
72°C - 45 sec.	

### Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unfilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

### Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	334-353	complement of 865-846
Fragment 2	638-656	complement of 1135-1116
Fragment 3	942-961	complement of 1422-1404
Fragment 4	1111-1130	complement of 1648-1629
Fragment 5	2051-2070	complement of 2570-2551
Fragment 6	2740-2759	complement of 3149-3130
Fragment 7	4051-4070	complement of 4409-4389
Fragment 8	4919-4938	complement of 5337-5316
Fragment 9	5289-5308	complement of 5645-5626
Fragment 10	6509-6528	complement of 6914-6895
Fragment 11	7331-7350	complement of 7740-7721
Fragment 12	7702-7721	complement of 8033-8014
Fragment 13	8576-8595	complement of 8905-8885
Fragment 14	9181-9200	complement of 9617-9598
Fragment 15	9605-9624	complement of 10064-10045
Fragment 16	10514-10533	complement of 10978-10959
Fragment 17	10758-10777	complement of 11151-11131
Fragment 18	11023-11041	complement of 11338-11319
Fragment 19	11377-11396	complement of 11869-11849
Fragment 20	11850-11871	complement of 12194-12175
Fragment 21	14659-14678	complement of 15021-15002
Fragment 22	14894-14913	complement of 15237-15218

### Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the NPR1 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the NPR1 Gene

Polymorphic Site Number	PolyId <sup>a</sup>	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	8199638	730	G	A		
PS2	8199735	731	G	C		
PS3	8199831	811	C	T	5	P2L
PS4	8199927	822	C	A	16	R6S
PS5	8203209	1235	G	C	429	A143A
PS6	8203306	1351	C	T	545	A182V
PS7	1568564	2184	T	C		
PS8	1568566	2472	A	G		
PS9	1568570	2979	G	C	1023	M341I
PS10	8214019	4345	T	A		
PS11	1568572	5290	T	C		
PS12	1568576	5537	G	A		
PS13	1568584	6900	G	A		
PS14	1568586	7410	A	T		
PS15	8206656	7947	C	T		
PS16	1568590	9313	G	A		
PS17	1568596	9619	G	A		
PS18	1568598	9675	A	T		
PS19	1568600	9904	C	T	2406	R802R
PS20	1568606	10004	G	A		
PS21	1568620	11062	C	T		

<sup>a</sup>PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

#### EXAMPLE 2

This example illustrates analysis of the NPR1 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4. Genotypes and Haplotype Pairs Observed for NPR1 Gene																	
Genotype Number	Polymorphic Sites															HAP Pair	
	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15		
1	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C	1	1
2	G	G	C	C	G	C	-	-	G	T	T	G	G	A/T	C	1	2
3	G	G	C	C	G	C	T	G	G	T	T	G	-	A	C/T	1	3
4	G	G/C	C	C	G/C	C/T	T	G	G	T/A	T/C	G	G	-	C	1	4
5	G	G	C	C	G	C	T/C	G	G	T	T/C	G/A	G/A	A	C	1	5
6	G/A	G	C/T	C/A	G	C	T	G	G	T	T	G	G	A	C	1	6
7	G	G	C	C	G	C	T	G/A	G	T	T	G	-	A	C	1	7
8	G	G	C	C	G	C	T	G	G	T	-	G/A	-	A	C	1	8
9	G	G	C	C	G	C	T/C	G	G	T	T/C	G/A	G/A	A/T	C	1	9
10	G	G	C	C	G	C	T	G	G	T	T/C	G	G	A	C	1	10
11	G	G	C	C	G	C	T	G	G/C	T	T	G	G	A	C	1	11
12	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C	1	12
13	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C	1	13
14	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C	1	14
15	G	G/C	C	C	G/C	C/T	T	G	G	T/A	T/C	G	G	A/T	C	2	4
16	G	G	C	C	G	C	T	G	G	T	T/C	G	G	A/T	C	2	10
17	G	G	C	C	G	C	T	G	G	T	T	G	G	A	T	3	3
18	G	G	C	C	G	C	T	G/A	G	T	-	G	-	A	C/T	3	7

Table 4.(contd) Genotypes and Haplotype Pairs Observed for NPR1 Gene

Genotype Number	Polymorphic Sites								
	PS 16	PS 17	PS 18	PS 19	PS 20	PS 21	HAP Pair		
1	G	G	A	C	G	C	1	1	
2	G	G	A	C	G	-	1	2	
3	G	G	A	C	G	C	1	3	
4	G	G/A	A/T	C/T	G	C	1	4	
5	G	G	A	C	G	C	1	5	
6	G	G/A	A	C	G	C	1	6	
7	G	G	A	C	G	C	1	7	
8	G	G	A	C	G	C	1	8	
9	G	G	A	C	G	C	1	9	
10	G	G	A	C	G	C	1	10	
11	G	G	A	C	G	C	1	11	
12	G	G	A	C	G	C/T	1	12	
13	G	-	A	C	G/A	C	1	13	
14	G/A	G	A	C	G	C	1	14	
15	G	G	A/T	C/T	G	C	2	4	
16	G	G	A	C	G	C	2	10	
17	G	G	A	C	G	C	3	3	
18	G	G	A	C	G	C	3	7	

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 14 human NPR1 haplotypes shown in Table 5 below.

Table 5. Haplotypes Identified in the NPR1 Gene															
Hap No.	Polymorphic Sites														
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	PS11	PS12	PS13	PS14	PS15
1	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C
2	G	G	C	C	G	C	T	G	G	T	T	G	G	T	C
3	G	G	C	C	G	C	T	G	G	T	T	G	G	A	T
4	G	C	C	C	C	T	T	G	G	A	C	G	G	A	C
5	G	G	C	C	G	C	C	G	G	T	C	A	A	A	C
6	A	G	T	A	G	C	T	G	G	T	T	G	G	A	C
7	G	G	C	C	G	C	T	A	G	T	T	G	G	A	C
8	G	G	C	C	G	C	T	G	G	T	T	A	G	A	C
9	G	G	C	C	G	C	C	G	G	T	C	A	A	T	C
10	G	G	C	C	G	C	T	G	G	T	C	G	G	A	C
11	G	G	C	C	G	C	T	G	C	T	T	G	G	A	C
12	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C
13	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C
14	G	G	C	C	G	C	T	G	G	T	T	G	G	A	C

Table 5(Contd.). Haplotypes Identified in the NPR1 Gene						
Hap No.	Polymorphic Sites					
	PS16	PS17	PS18	PS19	PS20	PS21
1	G	G	A	C	G	C
2	G	G	A	C	G	C
3	G	G	A	C	G	C
4	G	G	T	T	G	C
5	G	G	A	C	G	C
6	G	A	A	C	G	C
7	G	G	A	C	G	C
8	G	G	A	C	G	C
9	G	G	A	C	G	C
10	G	G	A	C	G	C
11	G	G	A	C	G	C
12	G	G	A	C	G	T
13	G	G	A	C	A	C
14	A	G	A	C	G	C

In Table 6 below, the number of chromosomes in unrelated individuals characterized by a given haplotype is shown, arranged by ethnic background of the subjects in the Index Repository. In Table 7 below, the number of unrelated subjects characterized by a given haplotype is shown, again arranged by ethnic background of the subjects in the Index Repository. In Tables 6 and 7, the following abbreviations are used: AF, African or African-American; AS, Asian; CA, Caucasian; HL, Hispanic-Latino; and AM, Native Americans.

**Table 6. Frequencies of Observed Haplotypes in Non-Related Individuals**

Hap No.	AF	AS	CA	HL	AM	Total
1	26	36	35	30	5	132
2	4	0	1	2	0	7
3	1	0	4	1	0	6
4	4	0	0	1	0	5
5	2	0	0	0	0	2
6	1	1	0	0	0	2
7	0	0	1	1	0	2
8	0	0	1	1	0	2
9	1	0	0	0	0	1
10	1	0	0	0	0	1
11	0	1	0	0	0	1
12	0	0	0	0	1	1
13	0	1	0	0	0	1
14	0	1	0	0	0	1

**Table 7. Frequencies of Observed Haplotype Pairs**

HAP Pair	AF	AS	CA	HL	AM	Total
1 1	6	16	16	13	2	53
2 1	4	0	1	1	0	6
3 1	1	0	2	1	0	4
3 3	0	0	1	0	0	1
4 1	4	0	0	0	0	4
4 2	0	0	0	1	0	1
5 1	2	0	0	0	0	2
6 1	1	1	0	0	0	2
7 1	0	0	1	1	0	2
7 3	0	0	0	0	0	0
8 1	0	0	1	1	0	2
9 1	1	0	0	0	0	1
10 1	5	0	0	0	0	5
10 2	1	0	0	0	0	1
11 1	0	1	0	0	0	1
12 1	0	0	0	0	1	1
13 1	0	1	0	0	0	1
14 1	0	1	0	0	0	1

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing

from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the natriuretic peptide receptor A/guanylate cyclase A (atriuretic peptide receptor A) (NPR1) gene of an individual which comprises determining whether the individual has one of the NPR1 haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4.
2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-21 on at least one copy of the individual's NPR1 gene.
3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-21 on both copies of the individual's NPR1 gene.
4. A method for genotyping the natriuretic peptide receptor A/guanylate cyclase A (atriuretic peptide receptor A) (NPR1) gene of an individual, comprising determining for the two copies of the NPR1 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21.
5. The method of claim 4, wherein the determining step comprises:
  - (a) isolating from the individual a nucleic acid mixture comprising both copies of the NPR1 gene, or a fragment thereof, that are present in the individual;
  - (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
  - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
  - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
  - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
6. The method of claim 4, which comprises determining for the two copies of the NPR1 gene present in the individual the identity of the nucleotide pair at each of PS1-21.
7. A method for haplotyping the natriuretic peptide receptor A/guanylate cyclase A (atriuretic peptide receptor A) (NPR1) gene of an individual which comprises determining, for one copy of the NPR1 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18,

PS19, PS20 and PS21.

8. The method of claim 7, wherein the determining step comprises:
  - (a) isolating from the individual a nucleic acid sample containing only one of the two copies of the NPR1 gene, or a fragment thereof, that is present in the individual;
  - (b) amplifying from the nucleic acid molecule a target region containing the selected polymorphic site;
  - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
  - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
  - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
9. A method for predicting a haplotype pair for the natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) gene of an individual comprising:
  - (a) identifying a NPR1 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21;
  - (b) enumerating all possible haplotype pairs which are consistent with the genotype;
  - (c) comparing the possible haplotype pairs to the data in Table 4; and
  - (d) assigning a haplotype pair to the individual that is consistent with the data.
10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-21.
11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-14 shown in Table 5 and the haplotype pair is selected from the haplotype pairs shown in Table 4, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.
12. The method of claim 11, wherein the trait is a clinical response to a drug targeting NPR1.
13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) gene at a polymorphic site selected from the group consisting of

- PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21.
14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the NPR1 gene at a region containing the polymorphic site.
  15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-24, the complements of SEQ ID NOS:4-24, and SEQ ID NOS:25-66.
  16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.
  17. The composition of claim 16, wherein the primer-extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:67-108.
  18. A kit for genotyping the NPR1 gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20 and PS21.
  19. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
    - (a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the natriuretic peptide receptor A/guanylate cyclase A (atriuretic peptide receptor A) (NPR1) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises a NPR1 isogene defined by a haplotype selected from the group consisting of haplotypes 1-14 in Table 5; and
    - (b.) a second nucleotide sequence which is complementary to the first nucleotide sequence.
  20. The isolated polynucleotide of claim 19, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
  21. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 19, wherein the organism expresses a NPR1 protein encoded by the first nucleotide sequence.
  22. The recombinant organism of claim 21, which is a nonhuman transgenic animal.
  23. The isolated polynucleotide of claim 19, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the NPR1 gene, the fragment comprising one or more polymorphisms selected from the group consisting of adenine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, cytosine at PS5, thymine at PS6, cytosine at PS7, guanine at PS8, cytosine at PS9, adenine at PS10, cytosine at PS11, adenine at PS12, adenine at PS13, thymine at PS14, thymine at PS15, adenine at PS16, adenine at PS17, thymine at PS18, thymine at PS19, adenine at PS20

and thymine at PS21.

24. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the NPR1 cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of a NPR1 isogene defined by one of the haplotypes shown in Table 5.
25. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 24, wherein the organism expresses a natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) protein encoded by the polymorphic variant sequence.
26. The recombinant organism of claim 25, which is a nonhuman transgenic animal.
27. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the NPR1 protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 5.
28. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 27.
29. A method for screening for drugs targeting the isolated polypeptide of claim 27 which comprises contacting the NPR1 polymorphic variant with a candidate agent and assaying for binding activity.
30. A computer system for storing and analyzing polymorphism data for the natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) gene, comprising:
  - (a) a central processing unit (CPU);
  - (b) a communication interface;
  - (c) a display device;
  - (d) an input device; and
  - (e) a database containing the polymorphism data;wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.
31. A genome anthology for the natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) gene which comprises NPR1 isogenes defined by any one of haplotypes 1-14 shown in Table 5.

1/11

## POLYMORPHISMS IN THE NPR1 GENE

GGATCCCAAA	CCAGCACACC	TTTCCCTCTT	CCCCCGAGGA	GACCAGGTAG	
GAGGCGAGGG	AAAAGGTGGG	GCGCAAGTGG	GCCCCGGTTG	CTTCCACACA	100
CACCCTCCGT	TCAGCCGTCC	TTTCCATCCC	GGCGAGGGCG	CACCTTCAGA	
GGGTCCCTGTC	CTCCAAAGAG	GTAGGCGTGG	GGCGGGCCGAG	ACCGGGGAAG	200
ATGGTCCACG	GGGAAGCGCG	CGGGCTGGGC	GGCGGGGAGG	AAGGAGTCTA	
TGATCCTGGA	TTGGCTCTTC	TGTCACTGAG	TCTGGGAGGG	GAAGCGGCTG	300
GGAGGGAGGG	TTCGGAGCTT	GGCTCGGGTC	CTCCACGGTT	CCCTCCGGAT	
AGCCGGAGAC	TTGGGCCCGC	CGGACGCCCC	TTCTGGCACA	CTCCCTGGGG	400
CAGGCGCTCA	CGCACGCTAC	AAACACACAC	TCCTCTTTCC	TCCCTCGCGC	
GCCCTCTCTC	ATCCTTCTTC	ACGAAGCGCT	CACTCGCACC	CTTTCTCTCT	500
CTCTCTCTCT	CTCTAACACG	CACGCACACT	CCCAGTTGTT	CACACTCGGG	
TCCTCTCCAG	CCCGACGTTT	TCCTGGCACC	CACCTGCTCC	GCGGCGCCCT	600
GCACGCCCCC	CTCGGTGCGC	CCCCTTGCGC	TCTCGGCCCA	GACCGTCGCA	
GCTACAGGGG	GCCTCGAGCC	CCGGGGTGAG	CGTCCCCGTC	CCGCTCCTGC	700
TCCTTCCCAT	AGGGACGCGC	CTGATGCCTG	GGACCGGCCG	CTGAGCCCAA	
		A C			
GGGGACCGAG	GAGGCCATGG	TAGGAGCGCT	CGCCTGCTGC	GGTGCCCGCT	800
GAGGCCATGC	CGGGGCCCCG	GCGCCCCGCT	GGCTCCCGCC	TGCGCCTGCT	
	T	A			
[exon 1: 807..					
CCTGCTCCTG	CTGCTGCCGC	CGCTGCTGCT	GCTGCTCCGG	GGCAGCCACG	900
CGGGCAACCT	GACGGTAGCC	GTGGTACTGC	CGCTGGCCAA	TACCTCGTAC	
CCCTGGTCGT	GGGCGCGCGT	GGGACCCGCC	GTGGAGCTGG	CCCTGGCCCA	1000
GGTGAAGGCG	CGCCCCGACT	TGCTGCCGGG	CTGGACGGTC	CGCACGGTGC	
TGGGCAGCAG	CGAAAACGCG	CTGGGCGTCT	GCTCCGACAC	CGCAGCGCCC	1100
CTGGCCGCGG	TGGACCTCAA	GTGGGAGCAC	AACCCCGCTG	TGTTCTTGGG	
CCCCGGCTGC	GTGTACGCCG	CCGCCCCAGT	GGGGCGCTTC	ACCGCGCACT	1200
GGCGGGTCCC	GCTGCTGACC	GCCGGCGCCC	CGGCGCTGGG	CTTCGGTGTC	
		C			
AAGGACGAGT	ATGCGCTGAC	CACCCGCGCG	GGGCCCAGCT	ACGCCAAGCT	1300
GGGGGACTTC	GTGGCGGCGC	TGCACCGACG	GCTGGGCTGG	GAGCGCCAAG	
CGCTCATGCT	CTACGCCTAC	CGGCCGGGTG	ACGAAGAGCA	CTGCTTCTTC	1400
	T				
CTCGTGGAGG	GGCTGTTTAT	GCGGGTCCGC	GACCGCCTCA	ATATTACGGT	
GGACCACCTG	GAGTTCGCCG	AGGACGACCT	CAGCCACTAC	ACCAGGCTGC	1500
TGCGGACCAT	GCCGCGCAAA	GGCCGAGGTG	AGACGCTGGC	ACACCCCGTC	
	..1527]				
CCGCCGCTTA	GCCGCAGGGC	CTCCCCCTCTG	ACCTGCCGGA	GGCATCGGGA	1600
CTTTCTCTCT	CATCTGGGGG	CACTCTTCTT	TCTCCTCGCC	GTTCTTCATT	
CTACTTTTCAG	CTCCCTGGCC	CTTTCTACAG	CTGAGTTTCT	ATTTCCCTCT	1700
CTTCTTCCGC	CACCCCCACC	ACGTCTCTAT	CCTCTCATCT	CCCCGACCCC	
CACCTATTCC	CTCCCACCCT	AGCACAGCTC	GGTTCGCGTC	CCTTTTTTCC	1800
TCCCACATTT	TCTCTCTTCC	CTATAGCCTT	CTCCCTTCTT	TCATCCTCTC	
CTCTCATGGC	GCCTCATCCC	CTCTCTTCTC	CCCCTCCCTC	TCCCTCCTCT	1900
CTCCCTCCTG	GCCCCATCCT	TCTCCACCTT	CAGCTCCACT	ATCCCCCTCT	
CCCTACCCGT	TCCTTCCTCC	CTTCCGCTC	CCCCTTCCTC	CTCCCGCCCA	2000
CCGCCCCGCA	CCCGCCCGTT	CCACCCTTCG	ACTTTCTCCT	GCTGTGGCCT	
AGGCTGAGCC	GGGAGTTACC	ACTTAACTCT	CACTGGGTCT	CTCCTGCACC	2100
CTATCTCTAA	ACTTCCTCCC	TTGGGTGCCC	CAGCTTTCCT	ACTCCTGTCT	

FIGURE 1A

					2/11	
CTCCCGCAGT	ACCTAGGCTT	CTCTCTCTGA	CTCTCCGTCT	TTCTCCAGTT	2200	
					C	
[exon 2: 2199..						
ATCTACATCT	GCAGCTCCCC	TGATGCCTTC	AGAACCCTCA	TGCTCCTGGC		
CCTGGAAGCT	GGCTTGTGTG	GGGAGGACTA	CGTTTTCTTC	CACCTGGATA	2300	
TCTTTGGGCA	AAGCCTGCAA	GGTGGACAGG	GCCCTGCTCC	CCGCAGGCCC		
TGGGAGAGAG	GGGATGGGCA	GGATGTCAGT	GCCCGCCAGG	CCTTTCAGGT	2400	
..2398]						
GAGTACCTAG	GTTTGAAGCC	CAGGCTGTCT	CAGCTTGTGG	CACATCATTT		
CTGGGCACTG	TGTCCCTCAG	CATCTGAAAG	AATTCCAGAA	AAGAGGTTTT	2500	
					G	
TGTCTGTTTG	TTTCTTTATG	CACTCCTGGT	AACTCACAGA	ACAGAAAAGA		
GGTTGGTGAT	GCTCACTGGG	AATTAGGCAA	TGAAGGGCAG	GGGACTGCCC	2600	
AGGGGCGCTT	CGCCACCAGC	AGGCTAAAAA	GATAAGAAAA	TGGGCTTGAG		
GCGGGAGGAG	GATAAAGTCC	CACAGCCTGG	ACAGGACTTG	GAGAAGGCAT	2700	
CCCATTTGGAT	CCCCTGCTTT	GGAATGGGCA	TCACTTCATG	CAGGGCATAG		
GGTCCAGTTT	GACCTTGAGC	TAAGCAGAGA	CGCAGCTCTG	GGAGGTGGGC	2800	
TCCCAACTGT	TGGGGCCCCA	CAGTACTAGG	GAATAGTCAG	CTCCCAACTC		
TCTGCTCTCC	ACTGACCCCT	TTCTCAGGCT	GCCAAAATCA	TTACATATAA	2900	
[exon 3: 2878..						
AGACCCAGAT	AATCCCGAGT	ACTTGGAATT	CCTGAAGCAG	TTAAAACACC		
TGGCCTATGA	GCAGTTCAAC	TTCACCATGG	AGGATGGCCT	GGTAAGAAGG	3000	
					C	
..2991]						
GGTCCCGGGA	CCCTCCAGCG	TGGACCTCCA	GCCCCCACTC	CATGACCCTC		
TGCCAGCCTC	CATCCTTCCC	TATTCCCAGT	TCTCCCCTTC	CTTCCCTCCC	3100	
TTCCCATTTG	TCCATGTTTC	TCGTGATGAT	GGAGGAGGAC	ACTGGCAAGT		
TCAGCCTCTG	AAACTCAGGT	CATCATCAGT	AATATGGAGA	CGATACATCC	3200	
TGCCCTGTCT	ACCTAGTAGG	ATTCAGGAAG	TGATGCTAAT	CCAAAGGCAT		
CGTTTAAATA	GTAAAATCTC	CCTGTGATAT	AGGGGTGTTA	TTTTCTCCCA	3300	
TCCTCTTCCA	AAATCCCAGT	GCCTCTTGTT	CCCTTCCCCA	CAGCTCCCAC		
CTCCATGCCC	TTCATATGCC	CACCCAGCC	GACCTCTGTT	TGCCCCTACA	3400	
GGTGAACACC	ATCCCAGCAT	CCTTCCACGA	CGGGCTCCTG	CTCTATATCC		
[exon 4: 3402..						
AGGCAGTGAC	GGAGACTCTG	GCACATGGGG	GAAGTGTAC	TGATGGGGAG	3500	
AACATCACTC	AGCGGATGTG	GAACCGAAGC	TTTCAAGGTC	AGGGCCTGGA		
..3537]						
GGTGGCTGGA	ATGGGCTGCC	TTGGGGGATG	AATCCCAGGT	GCCCAGTGTC	3600	
AAGCCATGAG	AAGCCTATTG	TCCTGCAGCA	GTTACCTATG	CACACCAGCC		
TTTTCTTCCA	CAGCTTTTTT	CAGGCCCATC	CCTCAGAAGT	CCTACAAAGT	3700	
GTCCAATCTC	AATCATCCCT	GCTGGGCACT	GAGTTCTTTT	ACCTTTCTTT		
TTCTTTTTTC	TTTTTTTTTT	GAGATGGAGT	CTCGCTCTGT	CCCCAAGACT	3800	
GGAGTGTTGG	GGTGCAATCT	CGGCTCACTT	CAACCTCCGC	CTCCCAGGTT		
CAAGCAATTC	TCCTGCCTCA	GCCTCCTGAG	TAGCTGGGAT	TACAGGTGCC	3900	
CTCCACCAAC	ACTTGGCTAA	TTTTTTTGAT	TTTTTTTAGT	AGAGACAGGG		
TTTACCACAG	TTGGTCAGGC	TGGTCTTGAA	CTCCTGACGT	CAGGTGATCT	4000	
GCCCGCCTCA	GCCTCCCAA	GTGCTGGGAT	TACAAGCATG	AGCCACAGTG		
CCCGGCCGTT	TTACCATTTA	CTATCATTTCT	GTATACATGT	ATGTTTGGA	4100	
GGCAAGGCAA	AAAAGATTAG	AGGATGAAGA	GATGAAGTGG	GGCACCCCTG		
AACTTCTATT	CTCTCAAACA	TAGTCATCTT	CCCCCATGTC	CTCAGGTGTG	4200	
[exon 5: 4196..						
ACAGGATACC	TGAAAATTGA	TAGCAGTGGC	GATCGGGAAA	CAGACTTCTC		

FIGURE 1B

3/11					
CCTCTGGGAT	ATGGATCCCG	AGAATGGTGC	CTTCAGGGTA	AGTTTGTGCA	4300
..4287]					
CCCAGAAGAC	AGTGCCAATT	CCAAATGACA	TCTCACCCCTC	CTACTTCCCC	
A					
CCCACAGCCC	TGCCAGGGCA	CCTGTTTATC	CTGTAGCCAT	TCCACCATGC	4400
CTGGACACTT	ACAAGAGCCC	TGGATAAAAC	AGACCCAGCT	CCAGTCTGGG	
GAAGCCACCA	GAATGATAGG	GACTCACAGG	CATCACACTT	GGGGAGCCCC	4500
ATGCCTGAGG	AGGGAGCACA	AGCCTGCCCT	CGGGGAGCTC	CGAAGGGAGG	
CAGGCAGGAC	CGCCTCCAG	CAGAGACAGG	GCTGTGAAAG	ATGCACATTA	4600
CACAGCTCTG	CAAGCGAGCA	GGGACAGGAA	GGCGCTGAGG	CCAATGGCCA	
CAAGGGACAG	GTCATCCAGA	GAAGGCCCTCC	TGGAAGACGG	GCACATGGAC	4700
TGGGCCCTGCG	AATGTAGGCT	AAGGTGAACA	TTACCTTCTC	CTGTTTTCTA	
CCAAGAAAAT	AAGTAGAGAA	AAATCAATGC	TTGGTTGGTA	CTTCAACCAA	4800
GATTATAAAC	TCCCTGAGTG	TAGAGATCGG	GTTCTAAATG	GAGTTTTCTT	
TATAAACCCC	TTGATAGTTT	TCAGGTGTTT	CCACTTGAGT	ACTATGTGTG	4900
TGGTATGAGG	TCCTGTGTCC	AGTTGCAGTG	GGGACTTGGT	AAGCAGGTGA	
CAACCCAGAT	ATATATGTAG	GCTCTAGAAG	CAGAGCTGGG	GTAGGTGGGA	5000
GGTGAGACTG	CTGCACTCAC	AGCATGCCTT	CCCCGCAGGC	CCTGGCCTAG	
CCACCACTCC	TGCTCTCCCT	TAGGTTGTAC	TGAACTACAA	TGGGACTTCC	5100
[exon 6: 5074..					
CAAGAGCTGG	TGGCTGTGTC	GGGGCGCAAA	CTGAACTGGC	CCCTGGGGTA	
CCCTCCTCCT	GACATCCCCA	AATGTGGCTT	TGACAACGAA	GACCCAGCAT	5200
GCAACCAAGG	TGACTGCCCC	TTGCCTTCCA	GGCCTCCATC	CCAGAGATGC	
..5209]					
TGCATCCTTC	CCCTAAGCAC	AGTCGAGTAG	GTGCTCCTGT	CCCATGCTGA	5300
C					
GGGCTTTCTG	GAGAATGACT	CCTGCCTTTT	TCTTCCCTTC	ATCCATCATC	
CCAGTTCCT	GATGGACTAT	TAGAAAGTTC	TTCTCCTGC	TGTCTAACCC	5400
AAATCTCTCT	TGCTGCAATA	TGGACTCTCT	CCTGCAGATC	ACCTTTCCAC	
[exon 7: 5438..					
CCTGGAGGTG	CTGGCTTTGG	TGGGCAGCCT	CTCCTTGCTC	GGCATTCTGA	5500
TTGTCTCCTT	CTTCATATAC	AGGTGAGCTG	TGATGTGGGG	GGTTGAGTGA	
A					
..5522]					
GGCTGGGGGA	CCCGGAGAAC	CAAGAGCAGA	GGAGGCGGTG	GGGACCCAGA	5600
GGGAAGAGGG	CAGGGGTGAA	GGGGCAGCAG	GGGAAAACCA	AGGGAGATGA	
GGAAGAAAGG	AGGCTTAAAA	GCCAGAGGAG	AAAGAAAGAG	AAGGGAATGG	5700
CAGGGCGAGG	GGAGGAGACA	AGGATAGGAA	TGGCCAAGGA	GAGTCAGAAA	
GATCCAAGAA	GCAGAGAAGT	TGATGGGTGA	CATCATAGGG	GCGTGGACTG	5800
GTTTTCTTGT	CTACTCTTGC	AGGCCAGATA	GGAAGCAACT	TTCTGAACCT	
TTGCAATCAT	GCCCATGTTA	GCTGAGGAGG	GTGAGCCCTG	GTGTGTGCCA	5900
GGTGCCCAAC	CTAGAATGGA	GAAGGGAGCT	GAATGAGCCT	TGTTCTTGCC	
GTCCAGTGGA	GGCTAAAATG	AAGTACAGGA	GGAGTTAATG	ATATACAAAA	6000
GCAAGGAGGG	AGGGGAGAAA	AATCACTGCT	GGTTGAGCAT	ATAATGTGTG	
CCAGGCACTT	CCACGTACAC	TATTTCTTTC	TTTCTTTTTT	TTTTTTTTTT	6100
TTTTTTTTTG	AGACGGAGTC	TCGCTCTGTT	GCCAGACTGG	AGTGCAGTGG	
CATGATCTAG	GCTCACTGCA	ACCTCCGCCT	CCCAGTTTCA	AGCAATTCTC	6200
CTGCCTCAGC	CTCCCATGTA	GCTGGGACTA	CAGGCACATG	CCACCACGCT	
CAGCTAATTT	TTGTATTTTT	AGTAGAGACA	GGGTTCACC	ATGTTGGCCA	6300
GGATGGTCTC	GATCTCTTGA	CCTCATGATC	CACCCACCTT	GGCCTCCCCA	
AGTGCTGGGA	TTACAGGCAT	GAGCCACTGT	GCCTGGCCTC	ATGTTCACTA	6400
TTTCTTTTCA	TTCTTATAAT	AGTTAAGAAT	GAAATAGATA	TTGCGGCCTC	

FIGURE 1C

4/11					
ATTCCCAAGT	AAGGACATTG	AGGTGATTCC	CCCAAGGTCC	CCAGTAAGGC	6500
AGAATTTCCC	CCAGCCATCC	TGATTCTCAG	TCCAGAGGAT	AGAATTTCCC	
CTCCATCTCT	GAGTGCATGG	TGTGGTCCCA	CGGCTCTGAG	GAGGGGCTGC	6600
TGAGCACCCCT	GCCCTGGGTC	AGCGGCTCAG	CCACAGGCTC	AGATGCAGCC	
TTCGTATCCC	AGGAAGATGC	AGCTGGAGAA	GGAAGTGGCC	TCGGAGCTGT	6700
[exon 8: 6663..					
GGCGGGTGCG	CTGGGAGGAC	GTTGAGCCCA	GTAGCCTTGA	GAGGCACCTG	
CGGAGTGCAG	GCAGCCGGCT	GACCCTGAGC	GGGGTAAGAA	CGCTGGTGTT	6800
..6783]					
TGTGTTGGGG	GGCAATAAAG	GAGAGGTGGG	TACAAGGGGC	AGTGCCTGAG	
GGATAGGTAA	GGGAGCAGGA	TTCTAGTCCC	AGCTCTGCTT	TCACTTGCTG	6900
A					
TGTGACCTTG	AGCGACTCAT	AGTCCCTCTC	CGAGACTGTC	TCAGATGATG	
ATTACAGCAG	CAGAGCCTCC	CTCACAGGGC	TCTTTTAAAG	GTCAGAGGAG	7000
ATAGTACCTG	TGAAAACACT	TTAAAAAATA	AAAAAGTAAA	TGAGGAGGAA	
ATTTTATGAT	GTGGAACATA	AAGCAGGGTG	GGCCAGGCAC	AGTGGCTCAC	7100
ATCTGCAATC	CCAGCACTTT	GGGAGACCGA	GGCAGGAGGA	TTGCTTGTGC	
CTGGGAGTTC	AAGACCAGCC	TGGGCAACAG	AGCAAGACAT	CGTCTCTACA	7200
AAGAATACAA	AGATTAGCAG	GGCATGGTGG	CGCATACTTG	TAGTCCCAGC	
TACTCTGGAG	GCTGAGGTGA	AAGGATCATC	TGAGCCCAGG	AGTCTGAGGC	7300
GGCAGTGACC	TAGGATAGCA	CCACTGCACT	CCAGCCTGGA	TGACACAATG	
ATACTACATC	TCAAAAAATA	ACCCAACAAC	AAAAAGGAAG	GGTGACACAA	7400
AGATAAGGCA	GGATAAGGCA	GGGAAATAAA	GACCAGAGCA	CAAGCAATCA	
T					
GGATGCAGAC	TGGGCCCACC	GGCTGACCAT	TCCTCCTGCT	CTCCCTCCTT	7500
TCAGAGAGGC	TCCAATTACG	GCTCCCTGCT	AACCACAGAG	GGCCAGTTCC	
[exon 9: 7505..					
AAGTCTTTGC	CAAGACAGCA	TATTATAAGG	TGGGCCTGGG	GAAAGATCAC	7600
..7579]					
TGGGCCTTGG	GACTGGGGCA	GGAGTGTAAT	CTGATGGAGG	ACTGGTGGGG	
GGTTCTGAGG	GAAGGAGTAA	GCTGGTGGGG	AGCAGCAGAT	GGGGGCCCTG	7700
GGGGTGGGCT	ATTGGGAACA	AGTGAGGGTC	CTGAGGGCAG	GGATGGGCTG	
TCGGGAGCAG	CTGGAATTCC	CAGGACATGG	GACCATGCTC	TTCACAGTGA	7800
CAGTCTCCAT	TCCATGCCCC	GGGCAACCTC	GTGGCTGTGA	AACGTGTGAA	
[exon 10: 7822..					
CCGTAAACGC	ATTGAGCTGA	CACGAAAAGT	CCTGTTTGAA	CTGAAGCATG	7900
..7899]					
TAATGTGGGG	AGTGAGGCAG	TGGCATGGAG	AAGGGGCCCT	CGGGGACGCA	
T					
AGGGGAGACTG	GCCAACAGAA	CTAGTTATGG	AGGGACCTCA	GGGTACCCCA	8000
AGAAAGGGGC	AGGGACTGGA	GCCCTGGATG	ACCTTCATCT	TGTGGTGGAG	
TGGGGGTATC	CTAAGTAGGA	GAAGAGACCA	CTGAGATAAC	CTGGAGGAAT	8100
CTTGAGGGGC	CATATGTGAT	GTCCCTGGGG	GAGAGAGGGC	TTAGGATGCC	
AGAGGGAGTA	GGAGCAGATT	CTGGGGAGGG	TGGGCTAAAG	GACATGGGTG	8200
GGAATCACCA	GGGAAGATCT	TAGTGATGGT	TGCAGAAAGT	GAATAAGGAG	
TTAAGAAGAG	TGAGGGTCCC	TGAAGCTAGT	GAGCAGCTTG	GTGAGGAGCG	8300
AGGTCTCTGT	CAAGCTCCTG	ATGCTGGTCC	CACTTGCAGA	TGCGGGATGT	
[exon 11: 8340..					
GCAGAATGAA	CACCTGACCA	GGTTTGTGGG	AGCCTGCACC	GACCCCCCCA	8400
ATATCTGCAT	CCTCACAGAG	TACTGTCCCC	GTGGGAGCCT	GCAGGTGAGG	
..8444]					
GGGACAAGGG	GTGTCAAGAA	ACCTGGGTTC	TAGCCCTGGC	TCTGCCCCTG	8500

FIGURE 1D

5/11				
ACTGGCCATA	AGACCCCAGG	CATGCCTCGC	CCTCTTTCTG	ACCTTTCTGG
CCCCATCTGT	AAAAATGGGA	GTTGGGGAAG	GGCAGTGGCA	CTAGAGTCAA
TCCAAAGTTT	TGTCCTGTTT	TACCAGTTCA	CATCAGTAGG	ACCCTGCACC
CTCCTCCAAC	TCCCAGGGGG	ATCTGCAGGG	GATTGGTCTT	GACTCTTATT
GCCCCAGCAG	GACATTCTGG	AGAATGAGAG	CATCACCTTG	GACTGGATGT
[exon 12: 8711..				
TCCGGTACTC	ACTCACCAAT	GACATCGTCA	AGGTATGCCC	CTAAGCACCT
..8782]				
ATTGGATGTG	TAGAGCAGGG	GCCAGGCATG	CTTCTCCTGG	CCACGGGTGT
AGGTCCCACT	CCTGGCCAAT	ACCTCTGCCC	ACTCACATTT	CCAGGGCATG
[exon 13: 8895..				
CTGTTTCTAC	ACAATGGGGC	TATCTGTTCC	CATGGGAACC	TCAAGTCATC
CAACTGCGTG	GTAGATGGGC	GCTTTGTGCT	CAAGATCACC	GACTATGGGC
TGGAGAGCTT	CAGGGACCTG	GACCCAGAGC	AAGGACACAC	CGTTTATGCC
AGTGAGCCTT	GACTCTTGAA	CCTAACACCT	GCCCCAGCA	CCACCCAGTA
..9051]				
GGGAGACTGA	TGCAAGGCCT	CTGATGGGCT	TGGGCATGCT	TGTCCTGACT
CCAGCCTCAA	TTCATTCAAC	CATGAAAAAG	GGAAGGCCAG	ACGAAGTGGT
TTCTAAGGCC	TCCTCTAGCT	CTAACACTCT	GTGATGCATC	CAGATCAGTT
TCGGCCACAC	CCTTGTTTCC	CCCTCACCCC	TTAGCTTTGG	GCTCCCTCAC
TCGGTGACTA	CCGACCTCTG	ACCCACAGAA	AAGCTGTGGA	CGGCCCTGA
A				
[exon 14: 9329..				
GCTCCTGCGA	ATGGCTTCAC	CCCCTGTGCG	GGGCTCCCAG	GCTGGTGACG
TATACAGCTT	TGGGATCATC	CTTCAGGAGA	TTGCCCTGAG	GAGTGGGGTC
TTCCACGTGG	AAGGTTTGGA	CCTGAGCCCC	AAAGGTGAGA	GGAGCACACC
..9484]				
TTCCTTAAAC	CCAGCCACAG	TCTCAACGAA	CCCCAGCCCC	AGGGAGAGGG
TCCCCTGGCA	GCACCACCAC	ACCTTCCTTC	TGTAATGGGG	TTCAGTCACC
ACCCTTTGAC	CCATTGCTGC	CAGTGACCAG	TCCCCCGCCC	CCATGCCTTG
A				
GTCTTGGA	TCCCCTGCCA	TCTCAGCTGG	TTGCCCCAGT	CTCTCACTAG
T				
GCCCTTGGCC	AGCCCCACCC	CTCAGCTCCT	CTACCCCCCA	ATACAGAGAT
[exon 15: 9747..				
CATCGAGCGG	GTGACTCGGG	GTGAGCAGCC	CCCCTTCCGG	CCCTCCCTGG
CCCTGCAGAG	TCACCTGGAG	GAGTTGGGGC	TGCTCATGCA	GCGGTGCTGG
GCTGAGGACC	CACAGGAGAG	GCCACCATTC	CAGCAGATCC	GCCTGACGTT
GCGCAAATTT	AACAGGTCCC	TGGTGTTTGT	CATGGATCCC	CCAGGCCCTT
T				
..9915]				
CCTCCACAGC	CACCATTTAC	CTAATGCTTC	TGGCTCTGGC	TTATCCCAGC
AGTGGCAGAG	GGAGACCACT	CACCTCCTCC	CTGTACATAG	TCAGCTCCAG
A				
CTCAGCACAG	CCTCATGACC	CTCTTCGCAA	GTACAGCATG	ACTCAGCTGT
CCCCACAGTC	CCCTGCCATT	CATGCCCTTT	CCCTCCACCA	TCGACACCCC
ACACCCTTCC	TGCCCACTCG	CCTTGCTGGC	CTCTAGACTT	CTCAGCAGTG
TGTAGGATAG	ATGGGCCTCC	CGCCTCCTGC	CCTGTAGGCT	CTTGGCCCTC
CACGGGAGCT	CCTGCCCCAC	CCCTTGATTT	CCCTTCCCCA	GCGTGCCCCAC
CAGGCCCAGT	TCCTCCAGAC	ACACCCTTCT	GTGGACATCA	CTTTGTCCGC
AATTGACCCT	TGTCATTCTC	CACCTCCTTT	ACCTCCTTCT	AACTCACTGG
GTTCAACAAA	GATGAACAAA	ATGTCCATAT	GTCTGAAGCT	TCATACTTGA

FIGURE 1E

6/11					
CCTTGGGGTTC	TCAGAAAAGA	ATTGAACTTT	CTTCCTTCTG	TTTTCCCCTG	10500
CTCCCCGGTA	TCCTGCTATG	CCCTCAACCC	TGAGCGTCTC	TAGAGACCTC	
ACTGCAGTCT	GGAGGGGGAA	GTGCCTAGGG	GCGGGCGCTC	ACGTAGGCTG	10600
TGCTGCTCCT	CTCTTACCAC	CCCCACCGCC	ACCCTCTGCC	CCCAGGGAGA	
[exon 16: 10646..					
ACAGCAGCAA	CATCCTGGAC	AACCTGCTGT	CCCGCATGGA	GCAGTACGCG	10700
AACAATCTGG	AGGAACTGGT	GGAGGAGCGG	ACCCAGGCAT	ACCTGGAGGA	
GAAGCGCAAG	GCTGAGGCCC	TGCTCTACCA	GATCCTGCCT	CAGTGAGTGC	10800
..10792]					
CTGAGTCTGG	GGACCCCCCC	CAACACAAAG	CCCCTGTCCC	GACCCCCAAC	
TCTGATCCTG	CACCTGCCCT	GACCCCTTAG	CTCAGTGGCT	GAGCAGCTGA	10900
[exon 17: 10881..					
AGCGTGGGGA	GACGGTGCAG	GCCGAAGCCT	TTGACAGTGT	TACCATCTAC	
TTCAGTGACA	TTGTGGGTTT	CACAGCGCTG	TCGGCGGAGA	GCACACCCAT	11000
GCAGGTAGGC	CAGGGTTCAG	CCACAGGTGC	CAGGCAAGCT	CAGCATCTGG	
..11004]					
ATCCCACCAG	ACCTGCCTTC	TGGTTCTGCT	TTACCCACCT	GACCCCAGGT	11100
T					
GGGGTCCCCT	ACTTCCTGTC	TCTCTTAGCT	TCTCTTCCCT	TCCAGGTGGT	
[exon 18: 11146..					
GACCCTGCTC	AATGACCTGT	ACACTTGCTT	TGATGCTGTC	ATAGACAACT	11200
TTGATGTGTA	CAAGGTGAGG	GTGGGAGTGG	GGATGGGAAG	GGACAGACAG	
..11214]					
ACATGGACAA	GGTCAGAAAA	AGATGAGGGG	TAGGCAGAAT	GATGTGGAGT	11300
CTTAAGAGAG	GAGATCGGGG	ACACGGGCAG	AGACAGTGAC	ACAGGGAGAC	
CCGGGAACAG	GCAGAGAACC	CATGTGGGAT	GGGGGATGAG	CAAAGACAGA	11400
TGAGGGTACA	GAATGACAGA	CGCTGCACCC	GGTGTGACGG	TGTGGCCGGC	
CGCACAGTTG	CAGCCGTCAA	GTCCTGCACC	CCCTCGCCAC	TCCCACAGGT	11500
[exon 19: 11499..					
GGAGACAATT	GGCGATGCCT	ACATGGTGGT	GTCAGGGCTC	CCTGTGCGGA	
ACGGGCGGCT	ACACGCCTGC	GAGGTAGCCC	GCATGGCCCT	GGCACTGCTG	11600
GATGCTGTGC	GCTCCTTCCG	AATCCGCCAC	CGGCCCCAGG	AGCAGCTGCG	
CTTGCGCATT	GGCATCCACA	CAGGTAAGGC	CACTGAAGGT	GCAGGCGGGC	11700
..11673]					
ATCCAGAGGC	CAAGGCTTTG	CAAGGGAAAC	TTGTCCCCTG	GCCCAGCCCC	
TCGCCCTTTC	ATCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	11800
GTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCACACACA	CACACACACA	
CACACACAGA	GCTGGGACCT	CAGATCCTGC	CTCCTGCCTG	TCTTGGATTG	11900
TCCACCTACC	TCCCTTAACA	CCCCTCCCTC	CCTCACTCGC	TGATGGGCTC	
TGCTCCTTCC	CTTGCTCCTC	CCAGGACCTG	TGTGTGCTGG	AGTGGTGGGA	12000
[exon 20: 11975..					
CTGAAGATGC	CCCGTTACTG	TCTCTTTGGG	GATACAGTCA	ACACAGCCTC	
AAGAATGGAG	TCTAATGGGG	AAGGTACAGT	GCCCCCTCCT	AGAGGGAATG	12100
..12073]					
GGGAGGGCAG	GGTGGCTGAG	GGAAATGCCA	TCCTGGGGCA	GCCTGTGCCT	
GCACAGCCCC	TTTCAGCTCC	TAGCCCTTTC	GCCTCCCAAG	TTCCCCTTCT	12200
CATAATATTA	AGAGTTCAAC	CTGGGCTCAT	CAACTTGACT	GTAACCAGAG	
ACTCAGGTTT	CTGCTGCCCC	TCTTGTCAAA	CGATGTAAAA	GTATTTCCGG	12300
GCCAGTGCTG	GAGAGTTCCC	AGCAGGAATC	TGATTTTAAG	ACCCTCTGTG	
GGCCGGGCGT	GGTGACTCAC	ACCTGTGATC	CCAGCACTTT	GGGAAGCTGA	12400
GGCAGGCGGA	TCACCTGAGG	TCGGGGGTTT	CGAGACCAGC	CTGACCAACA	
TGATGAAATC	CCGTCTCTAC	TAAAAATACA	AAAAACTAGC	CAGGTGTGAT	12500

FIGURE 1F

7/11.					
GGCAGGCTCC	TGTAATCCCA	GCTACTTGGG	AGGCTTGAGG	CAGAAGAATT	
GCTTGAACCC	GGGAGGCAGA	GGTTGCGATG	AGCCAAGATT	ACACCACGCA	12600
CCCCAGCTTG	GGCAATAAGA	GTTAAACTCT	GTCTCAAAAA	AAAAAAAAAA	
AAAAAAAAAA	AGGGCCCTCT	GCTCCACCTT	TGATGTGGTA	AAGATGGCTT	12700
CAGAGCCAGC	ATAAGTGAGG	CTGTGAATCT	CAGCTCCACA	GCTGGCTGTG	
TGTCAGTTTG	CTATACCTCT	CTGAGCCATG	GTTTTCCTCA	TCTGTAAAAA	12800
GAGGGAAAAA	ATCTATCTCA	CAGGAATTAT	GTGAGAAACC	CATTAAAAAT	
GTCTACCACA	TAATTGTCAT	TTAACTTTTC	CAAGCCTTAG	CGGATTATCT	12900
GTAAAAATGAT	GTCTATCTCA	GGATTGCAAG	AAGCCTAGCA	CAAACCCTGG	
TACCCAGCAG	GCACCTAATA	AATTCTTACT	CCTACCCGCC	CCTTGCTCTT	13000
GCCTCCTGTT	TATCTTCTAT	CCTTCTGCTG	TATTCGACAC	AATTCAATGC	
AGTAAACATT	TATTGAGTGA	CTACTGAGTG	CCAGGCCCTG	GGATAGTAAC	13100
ATGGCCCAGA	TCCAGAGTTA	GCTGAGAAAT	TCATGTGGAC	CCCATCTAAA	
CCTTATGGTG	AAAGAAAGGC	TGCTTGGGAG	CCAGTCCTGG	GAGCCCAGAG	13200
GGATCTAGTT	CGGCAAATAT	TCCCTGGGCA	CTATTTGGGG	GCTGCAGAGT	
CAGCCCTTGT	TGAGGGTCCA	GTCTCAAGG	AGCACATTCC	CAGAAATGTT	13300
CACATTCTGG	CGCTGGGGTG	CTGTAATCCC	AGCACTTTGG	GAGGCCGAGG	
TGGGCAGATC	ACTTGAGGCC	AGGAGTGGAG	ACTAGCCTGG	CCAACATGGT	13400
GACCTCCTGT	CTCTACTAAA	AATACAAAAA	ATTAGCTGGG	CGTGGTGGCA	
CGTGCCCGTA	ATCCCAGCTA	CTCAGGAGGC	TTGAGACATG	AAAATCACTT	13500
GAACCCAGGA	GGTGGATGTT	GCAGTGAGCC	GAGACTGCAC	CCCTGGGCAA	
CAGAGCGAGA	CTCTGTCTCA	AAAAAAAAAA	AGAGAGAAAG	AAAGAAAAGA	13600
AAAGAAAGAA	ACTGTTAAAC	ACAACAAGGC	CACTGTGATT	GATGCAAACC	
CCAGAAGTAG	GGACATGAGT	TCAGACAGTG	GTCAAAGAGA	GGGTGTGGCA	13700
ATATTGGGCC	CCACTCCATC	ACTGACCTCC	TCAGCCACTT	GGGCAGATCA	
CCCTGGGCCT	CAGTTCCTCG	GCCACAAAAT	GAGGGTATAG	CATGAAATCA	13800
TGAAAGCAAC	AATTTACATA	GTGCTTCCTA	GGTAGCACAT	TCCGTTTGAA	
TACTTTATGG	ATGTTAAATT	TAATCCTCAC	AACAAGGTTT	TGAGATGGGT	13900
ACTGACACTA	TCAGCATTTT	ACAGATTAGG	AAAATGAAGC	AGAGAGAATT	
TATTTTACAT	ACCTAAGCAA	GTATCCAAGC	TGAGGTTTCAT	ACTGAGGCAG	14000
TGCAGGATCC	AAAGTGCCAG	CTCCTAACCA	CCATGCTGTG	TAGAGCCGGG	
TGACACTCCA	GAGAGTGCTG	TCCAACAGGA	TGTTCCATAG	TCATGAAAAT	14100
GTTCTGTATT	CTGTGCTGTC	CAATACAGTA	GCCTCTAGGC	ACATATGGCT	
ACTTATCACT	GGAAATGTGA	CGGGTGCAAC	TGAGGCCCTG	ATTTTTTTTT	14200
TTTTTTTGGG	GACAGAGTTT	CGCTCTGTCT	CCCAGCCTGG	ATGGAGTGCA	
GTGGTGCAAT	CTCGGCTCAC	TGCAACCTCC	GCCTCCCAGG	TTCAAGCGAT	14300
TCTCCTGCCT	CAGCCTCCCA	AGTAGCTGGA	ATTACAGGTG	AGTGCCACCA	
CACACAGCTA	ATTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	CGCCATATTG	14400
GCCAGGATGG	TCTCGAACTC	CTGGCCTCAA	GTGATCCTCC	TGCCTCAGCC	
TCCCAAAGTG	CTGGGATTAC	AGGTGTGAGC	CACAGCACCC	AGCCTGAATT	14500
TTTAACTGTA	TTTAGTTTAA	ATTAATTTAA	GTTGAAACAG	GCACATGTGA	
TTAGTGGCTA	CTGTATTGGA	TTACACAGCT	CCAGAGTTCT	AAATGAGAGG	14600
CTAATGTGGT	CACGCACTAC	ATTCAGGGGG	TGGGGCCCCCT	CTGAGCTAGA	
GGGCTTCCTG	GCCCAAAGA	GGGAGAGAGG	GTACCTGTCC	ACCTGTCCAC	14700
CCCCACAGTC	CTGGTCTCT	TTTGCCTCTA	CTTTCCTGCT	CTCCTCTCTC	
ACATTGCTCA	CCTTCCCTTC	TCCCCTGTCC	TACCCAGCCC	TGAAGATCCA	14800
[exon 21: 14788..					
CTTGTCTTCT	GAGACCAAGG	CTGTCCTGGA	GGAGTTTGGT	GGTTTTCGAGC	
TGGAGCTTCG	AGGGGATGTA	GAAATGAAGG	TAGAGCGAGA	AGCCTCTGCC	14900
..14879]					
CTCCCCACCT	TTTGGGGTCC	TAGAGGGAGT	TACCCTTCTC	AAGCAGCCGA	
TGCCACTCCC	ATCCCTAAGG	CTCTCATCTG	ACTGGGGAAA	GGGCATGTGC	15000

FIGURE 1G

8/11

CACTCCCCAG	CCCATCCTCT	TTTTTCCCTC	CAGGGCAAAG	GCAAGGTTCTG	
[exon 22: 15034..					
GACCTACTGG	CTCCTTGGGG	AGAGGGGGAG	TAGCACCCGA	GGCTGACCTG	15100
..15096]					
CCTCCTCTCC	TATCCCTCCA	CACCTCCCCT	ACCCTGTGCC	AGAAGCAACA	
GAGGTGCCAG	GCCTCAGCCT	CACCCACAGC	AGCCCCATCG	CCAAAGGATG	15200
GAAGTAATTT	GAATAGCTCA	GGTGTGCTGA	CCCCAGTGAA	GACACCAGAT	
AGGACCTCTG	AGAGGGGACT	GGCATGGGGG	GATCTCAGAG	CTTACAGGCT	15300
GAGCCAAGCC	CACGGCCATG	CACAGGGACA	CTCACACAGG	CACACGCACC	
TGCTCTCCAC	CTGGACTCAG	GCCGGGCTGG	GCTGTGGATT	CCTGATCCCC	15400
TCCCCTCCCC	ATGCTCTCCT	CCCTCAGCCT	TGCTACCCTG	TGACTTACTG	
GGAGGAGAAA	GAGTCACCTG	AAGGGGAACA	TGAAAAGAGA	CTAGGTGAAG	15500
AGAGGGCAGG	GGAGCCCACA	TCTGGGGCTG	GCCACAATA	CCTGCTCCCC	
CGACCCCCTC	CACCCAGCAG	TAGACACAGT	GCACAGGGGA	GAAGAGGGGT	15600
GGCGCAGAAG	GGTTGGGGGC	CTGTATGCCT	TGCTTCTACC	ATGAGCAGAG	
ACAATTAAAA	TCTTTATTCC	AGTGACAGTG	TCTCTTCTTG	AGGGAGAGAG	15700
GGTTGCCAGA	AAACAGTCAG	TTCTCCACTC	TCTACTTCAA	ATAAGACTCA	
CTTCTTG TTC	TACAAGGGTC	TAGAAGGAAA	AGTAAAAAAA	AAAGACTCTC	15800
GATTCTTAAC					15810

FIGURE 1H

9/11

## POLYMORPHISMS IN THE CODING SEQUENCE OF NPR1

ATGCCGGGGC	CCCGGCGCCC	CGCTGGCTCC	CGCCTGCGCC	TGCTCCTGCT	
T	A				
CCTGCTGCTG	CCGCCGCTGC	TGCTGCTGCT	CCGGGGCAGC	CACGCGGGCA	100
ACCTGACGGT	AGCCGTGGTA	CTGCCGCTGG	CCAATACCTC	GTACCCCTGG	
TCGTGGGCGC	GCGTGGGACC	CGCCGTGGAG	CTGGCCCTGG	CCCAGGTGAA	200
GGCGCGCCCC	GACTTGCTGC	CGGGCTGGAC	GGTCCGCACG	GTGCTGGGCA	
GCAGCGAAAA	CGCGCTGGGC	GTCTGCTCCG	ACACCGCAGC	GCCCCTGGCC	300
GCGGTGGACC	TCAAGTGGGA	GCACAACCCC	GCTGTGTTCC	TGGGCCCCGG	
CTGCGTGTAC	GCCGCCGCCC	CAGTGGGGCG	CTTCACCGCG	CACTGGCGGG	400
TCCCCGCTGCT	GACCGCCGGC	GCCCCGGCGC	TGGGCTTCGG	TGTCAAGGAC	
		C			
GAGTATGCGC	TGACCACCCG	CGCGGGGCCC	AGCTACGCCA	AGCTGGGGGA	500
CTTCGTGGCG	GCGCTGCACC	GACGGCTGGG	CTGGGAGCGC	CAAGCGCTCA	
			T		
TGCTCTACGC	CTACCGGCCG	GGTGACGAAG	AGCACTGCTT	CTTCCTCGTG	600
GAGGGGCTGT	TCATGCGGGT	CCGCGACCGC	CTCAATATTA	CGGTGGACCA	
CCTGGAGTTC	GCCGAGGACG	ACCTCAGCCA	CTACACCAGG	CTGCTGCGGA	700
CCATGCCGCG	CAAAGGCCGA	GTTATCTACA	TCTGCAGCTC	CCCTGATGCC	
TTCAGAACCC	TCATGCTCCT	GGCCCTGGAA	GCTGGCTTGT	GTGGGGAGGA	800
CTACGTTTTT	TTCCACCTGG	ATATCTTTGG	GCAAAGCCTG	CAAGGTGGAC	
AGGGCCCTGC	TCCCCGCAGG	CCCTGGGAGA	GAGGGGATGG	GCAGGATGTC	900
AGTGCCCGCC	AGGCCTTTCA	GGCTGCCAAA	ATCATTACAT	ATAAAGACCC	
AGATAATCCC	GAGTACTTGG	AATTCTTGAA	GCAGTTAAAA	CACCTGGCCT	1000
ATGAGCAGTT	CAACTTCACC	ATGGAGGATG	GCCTGGTGAA	CACCATCCCA	
		C			
GCATCCTTCC	ACGACGGGCT	CCTGCTCTAT	ATCCAGGCAG	TGACGGAGAC	1100
TCTGGCACAT	GGGGGAACTG	TTACTGATGG	GGAGAACATC	ACTCAGCGGA	
TGTGGAACCG	AAGCTTTCAA	GGTGTGACAG	GATACCTGAA	AATTGATAGC	1200
AGTGGCGATC	GGGAAACAGA	CTTCTCCCTC	TGGGATATGG	ATCCCGAGAA	
TGGTGCCTTC	AGGGTTGTAC	TGAACTACAA	TGGGACTTCC	CAAGAGCTGG	1300
TGGCTGTGTC	GGGGCGCAAA	CTGAACTGGC	CCCTGGGGTA	CCCTCCTCCT	
GACATCCCCA	AATGTGGCTT	TGACAACGAA	GACCCAGCAT	GCAACCAAGA	1400
TCACCTTTCC	ACCCTGGAGG	TGCTGGCTTT	GGTGGGCAGC	CTCTCCTTGC	
TCGGCATTCT	GATTGTCTCC	TTCTTCATAT	ACAGGAAGAT	GCAGCTGGAG	1500
AAGGAACTGG	CCTCGGAGCT	GTGGCGGGTG	CGCTGGGAGG	ACGTTGAGCC	
CAGTAGCCTT	GAGAGGCACC	TGCGGAGTGC	AGGCAGCCGG	CTGACCCTGA	1600
GCGGGAGAGG	CTCCAATTAC	GGCTCCCTGC	TAACCACAGA	GGGCCAGTTC	
CAAGTCTTTG	CCAAGACAGC	ATATTATAAG	GGCAACCTCG	TGGCTGTGAA	1700
ACGTGTGAAC	CGTAAACGCA	TTGAGCTGAC	ACGAAAAGTC	CTGTTTGAAC	
TGAAGCATAT	GCGGGATGTG	CAGAATGAAC	ACCTGACCAG	GTTTGTGGGA	1800
GCCTGCACCG	ACCCCCCAA	TATCTGCATC	CTCACAGAGT	ACTGTCCCCG	
TGGGAGCCTG	CAGGACATTC	TGGAGAATGA	GAGCATCACC	CTGGACTGGA	1900
TGTTCCGGTA	CTCACTCACC	AATGACATCG	TCAAGGGCAT	GCTGTTTCTA	
CACAATGGGG	CTATCTGTTT	CCATGGGAAC	CTCAAGTCAT	CCAAGTGCCT	2000
GGTAGATGGG	CGCTTTGTTC	TCAAGATCAC	CGACTATGGG	CTGGAGAGCT	
TCAGGGACCT	GGACCCAGAG	CAAGGACACA	CCGTTTATGC	CAAAAAGCTG	2100
TGGACGGCCC	CTGAGCTCCT	GCGAATGGCT	TCACCCCTG	TGCGGGGCTC	
CCAGGCTGGT	GACGTATACA	GCTTTGGGAT	CATCCTTCAG	GAGATTGCCC	2200

FIGURE 2A

10/11

TGAGGAGTGG	GGTCTTCCAC	GTGGAAGGTT	TGGACCTGAG	CCCCAAAGAG	
ATCATCGAGC	GGGTGACTCG	GGGTGAGCAG	CCCCCCTTCC	GGCCCTCCCT	2300
GGCCCTGCAG	AGTCACCTGG	AGGAGTTGGG	GCTGCTCATG	CAGCGGTGCT	
GGGCTGAGGA	CCCACAGGAG	AGGCCACCAT	TCCAGCAGAT	CCGCCTGACG	2400
TTGCGCAAAT	TTAACAGGGA	GAACAGCAGC	AACATCCTGG	ACAACCTGCT	
T					
GTCCCGCATG	GAGCAGTACG	CGAACAATCT	GGAGGAACTG	GTGGAGGAGC	2500
GGACCCAGGC	ATACCTGGAG	GAGAAGCGCA	AGGCTGAGGC	CCTGCTCTAC	
CAGATCCTGC	CTCACTCAGT	GGCTGAGCAG	CTGAAGCGTG	GGGAGACGGT	2600
GCAGGCCGAA	GCCTTTGACA	GTGTTACCAT	CTACTTCAGT	GACATTGTGG	
GTTTCACAGC	GCTGTGCGCG	GAGAGCACAC	CCATGCAGGT	GGTGACCCTG	2700
CTCAATGACC	TGTACACTTG	CTTTGATGCT	GTCATAGACA	ACTTTGATGT	
GTACAAGGTG	GAGACAATTG	GCGATGCCTA	CATGGTGGTG	TCAGGGCTCC	2800
CTGTGCGGAA	CGGGCGGCTA	CACGCCTGCG	AGGTAGCCCG	CATGGCCCTG	
GCACTGCTGG	ATGCTGTGCG	CTCCTTCCGA	ATCCGCCACC	GGCCCCAGGA	2900
GCAGCTGCGC	TTGCGCATTG	GCATCCACAC	AGGACCTGTG	TGTGCTGGAG	
TGGTGGGACT	GAAGATGCCC	CGTTACTGTC	TCTTTGGGGA	TACAGTCAAC	3000
ACAGCCTCAA	GAATGGAGTC	TAATGGGGAA	GCCCTGAAGA	TCCACTTGTC	
TTCTGAGACC	AAGGCTGTCC	TGGAGGAGTT	TGGTGGTTTC	GAGCTGGAGC	3100
TTCGAGGGGA	TGTAGAAATG	AAGGGCAAAG	GCAAGGTTTCG	GACCTACTGG	
CTCCTTGGGG	AGAGGGGGAG	TAGCACCCGA	GGCTGA		3186

FIGURE 2B

11/11

## ISOFORMS OF THE NPR1 PROTEIN

MPGPRRPAGS	RLRLLLLLLL	PPLLLLLLRGS	HAGNLTAVV	LPLANTSYPW	
L S					
SWARVGPAVE	LALAQVKARP	DLLPGWTVRT	VLGSSSENALG	VCSDTAAPLA	100
AVDLKWEHNP	AVFLGPGCVY	AAAPVGRFTA	HWRVPLLTAG	APALGFGVKD	
EYALTTRAGP	SYAKLGDFVA	ALHRRLGWER	QALMLYAYRP	GDEEHCFFLV	200
		V			
EGLFMRVRDR	LNITVDHLEF	AEDDLSHYTR	LLRTMPRKGR	VIYICSSPDA	
FRTLMLLALE	AGLCGEDYVF	FHLDIFGQSL	QGGQGPAPRR	PWERGDGQDV	300
SARQAFQAAK	IITYKDPDNP	EYLEFLKQLK	HLAYEQFNFT	MEDGLVNTIP	
			I		
ASFHDGLLLY	IQAVTETLAH	GGTVTDGENI	TQRMWNRSFQ	GVTGYLKIDS	400
SGDRETDFSL	WDMDPENGAF	RVVLNNGTS	QELVAVSGRK	LNWPLGYPPP	
DIPKCGFDNE	DPACNQDHLS	TLEVLALVGS	LSLLGILIVS	FFIYRKMQL	500
KELASELWRV	RWEDVEPSSL	ERHLRSAGSR	LTLSGRGSNY	GSLLTTEGQF	
QVFAKTAYYK	GNLVAVKRVN	RKRIELTRKV	LFELKHMRDV	QNEHLTRFVG	600
ACTDPPNICI	LTEYCPRGSL	QDILENESIT	LDWMFRYSLT	NDIVKGMLFL	
HNGAICSHGN	LKSSNCVVDG	RFVLKITDYG	LESFRDLDP	QGHTVYAKKL	700
WTAPELLRMA	SPPVRGSQAG	DVYSFGIILQ	EIALRSGVFH	VEGLDLSPKE	
IIERVTRGEQ	PPFRPSLALQ	SHLEELGLLM	QRCWAEDPQE	RPPFQQIRLT	800
LRKFNRENS	NILDNLLSRM	EQYANNLEEL	VEERTQAYLE	EKRKAEALLY	
QILPHSVAEQ	LKRGETVQAE	AFDSVTIYFS	DIVGFTALSA	ESTPMQVVTL	900
LNDLYTCFDA	VIDNFDVYKV	ETIGDAYMVV	SGLPVRNGRL	HACEVARMAL	
ALLDAVRSFR	IRHRPQEQLR	LRIGIHTGPV	CAGVVGLKMP	RYCLFGDTVN	1000
TASRMESNGE	ALKIHLSET	KAVLEEFGGF	ELELRGDVEM	KGKGKVRTYW	
LLGERGSSTR	G				1061

FIGURE 3

## SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.  
 Bentivegna, Steven C.  
 Choi, Julie Y.  
 Kliem, Stefanie E.  
 Nandabalan, Krishnan

<120> Haplotypes of the NPR1 Gene

<130> MWH-0346PCT NPR1

<140> To be assigned

<141> 2001-04-16

<150> 60/197,330

<151> 2000-04-14

<160> 109

<170> PatentIn Ver. 2.1

<210> 1

<211> 15810

<212> DNA

<213> Homo sapiens

<400> 1

```

ggatcccaaa ccagcacacc ttccctctt ccccgagga gaccaggtag gaggcgaggg 60
aaaaggtggg ggcgaagtgg gcccggttg ctccacaca caccctcgt tcagccgtcc 120
tttccatccc ggcgagggcg cacttcaga gggctctgtc ctcaaagag gtaggcgtgg 180
ggcggccgag accggggaag atgggtccacg gggaagcgcg cgggctgggc ggcggggagg 240
aaggagtcta tgatcctgga ttggctcttc tgtcactgag tctgggaggg gaagcggtg 300
ggagggaggg ttcggagctt ggctcgggtc ctccacggtt ccctcggat agccggagac 360
ttgggccggc cggacgcccc ttctggcaca ctccctgggg caggcgctca cgcacgtac 420
aaacacacac tcctctttcc tcctcgcgc gccctctctc atccttcttc acgaagcgct 480
cactcgcacc ctttctctct ctctctctct ctctaacacg cacgcacact ccagttggt 540
cacactcggg tcctctccag ccgacgttc tcctggcacc cactgctcc gcggcgccct 600
gcacgcccc ctcggtcgcg ccccttgcg tcctggccca gaccgtcgca gctacagggg 660
gcctcgagcc ccggggtgag cgtccccgtc ccgtcctgc tccttccat agggacgcgc 720
ctgatgcctg ggaccggccg ctgagcccaa ggggaccgag gaggccatgg taggagcgct 780
cgctgctgc ggtgcccgt gaggccatgc cggggccccg gcgccccgt ggctcccgcc 840
tgcgctgct cctgctcctg ctgctgccgc cgctgctgct gctgctccgg ggcagccacg 900
cgggcaacct gacggtagcc gtgggtactgc cgctggccaa tacctcgtac ccctggctgt 960
gggcgcgcgt gggaccgcgc gtggagctgg ccctggccca ggtgaaggcg cgcgccgact 1020
tgctgccggg ctggacggtc cgcacggtgc tgggcagcag cgaacacgcg ctgggcgtct 1080
gctccgacac cgcagcgcgc ctggccgcgg tggacctcaa gtgggagcac aaccccgctg 1140
tgttcttggg ccccggtgct gtgtacgccg ccgccccagt ggggcgcttc accgcgcact 1200

```

```

ggcggtcccc gctgctgacc gccggcgccc cggcgctggg cttcggtgtc aaggacgagt 1260
atgcgctgac caccgcgcgc gggcccagct acgccaagct gggggacttc gtggcgggcg 1320
tgcaccgacg gctgggctgg gagcgccaag cgctcatgct ctacgcctac cggccgggtg 1380
acgaagagca ctgctttctt ctcgtggagg ggctgttcat gcggttcgcg gaccgcctca 1440
atattacggt ggaccacctg gagttcgccg aggacgacct cagccactac accaggctgc 1500
tgcggaacct gccgcgcaaa ggccgaggtg agacgctggc acaccccgtc ccgcgcctta 1560
gccgcagggc ctccctctct acctgccgga ggcatcgga ctttctctct catctggggg 1620
cactcttctt tctcctcgcc gttcttcatt ctactttcag ctccctggcc ctttctacag 1680
ctgagtttct atttccctct cttcttccgc cacccccacc acgtctctat cctctcatct 1740
ccccgacccc cactcattcc ctcccaccct agcacagctc ggttcgggtc cttttttccc 1800
tcccacattt tctctcttcc ctatagcctt ctcccttctt tcctcctctc ctctcatggc 1860
gcctcatccc ctctcttctc cccctccctc tcctcctct ctcctcctg gccccatcct 1920
tctccacctt cagctccact atccccctct ccctaccgt tccttctct cttccgcctc 1980
cccccttctc ctccgcgcc cgcgccgca ccgcgccgt ccacccttg actttctct 2040
gctgtggcct aggtgagcc gggagttacc acttaactct cactgggtct ctctgcacc 2100
ctatctctaa acttctctcc ttgggtgccc cagctttcct actcctgtct ctccgcagt 2160
acctaggctt ctctctctga ctctcgtct ttctccagtt atctacatct gcagctcccc 2220
tgatgccttc agaaccctca tgctcctggc cctggaagct ggcttgtgtg gggaggacta 2280
cgttttcttc cacctggata tctttgggca aagcctgcaa ggtggacagg gccctgctcc 2340
ccgcaggccc tgggagagag gggatgggca ggatgtcagt gcccgccagg ctttccaggt 2400
gagtacctag gtttgaagcc caggctgtct cagcttgtgg cacatcattt ctgggcaactg 2460
tgtccctcag catctgaaag aattccagaa aagaggtttt tgtctgtttg tttctttatg 2520
cactcctggt aactcacaga acagaaaaga ggttggtgat gctcactggg aattaggcaa 2580
tgaagggcag gggactgcc aggggcgctt cgccaccagc aggctaaaaa gataagaaaa 2640
tgggcttgag gcgggaggag gataaagtcc cacagcctgg acaggacttg gagaaggcat 2700
cccattggat cccctgcttt ggaatgggca tcacttcagt cagggcatag ggtccagttt 2760
gaccttgagc taagcagaga cgcagctctg ggaggtgggc tcccaactgt tggggcccca 2820
cagtactagg gaatagtcag ctcccactc tctgctctcc actgaccctt ttctcaggt 2880
gccaaaatca ttacatataa agaccagat aatcccagat acttgggaatt cctgaagcag 2940
ttaaaacacc tggcctatga gcagttcaac ttcaccatgg aggatggcct ggtaagaagg 3000
ggtcccgga cctccagcg tggacctca gccccactc catgaccctc tgcagcctc 3060
catccttccc tattccagt tctcccttc cttccctccc ttcctattgt tccatgtttc 3120
tcgtgatgat ggaggaggac actggcaagt tcagcctctg aaactcaggt catcatcagt 3180
aatatggaga cgatacatcc tgccctgtct acctagtagg attcaggaag tgatgctaata 3240
ccaaaggcat cgtttaaata gtaaaatctc cctgtgatat aggggtgtta ttttctccca 3300
tctcttccca aaatccagt gcctcttggt ccttcccca cagctcccac ctccatgcc 3360
ttcatatgcc caccagcc gacctctgtt tgcccctaca ggtgaacacc atcccagcat 3420
ccttccacga cgggctcctg ctctatatcc aggcagtgc ggagactctg gcacatgggg 3480
gaactgttac tgatggggag aacatcactc agcggatgtg gaaccgaagc tttcaagggtc 3540
agggcctgga ggtggetgga atgggctgcc ttgggggatg aatcccaggt gccagtgctc 3600
aagccatgag aagcctattg tctgcagca gttacctatg cacaccagcc ttttctcca 3660
cagctttttt caggcccatc cctcagaagt cctacaaagt gtccaatctc aatcatccct 3720
gctgggcact gagttctttt acctttcttt ttctttttt ttttttttt gagatggagt 3780
ctcgtctgtt ccccaagact ggagtgtggt ggtgcaatct cggctcactt caacctcgc 3840
ctcccaggtt caagcaattc tctgcctca gcctcctgag tagctgggat tacagggtgc 3900
ctccaccaac acttggtctaa ttttttgtat ttttttagt agagacaggg tttcaccacg 3960
ttggtcaggc tggctctgaa ctctgacgt caggtgatct gccgcctca gcctcccaaa 4020
gtgctgggat tacaagcatg agccacagt cccggccgtt ttaccattta ctatcattct 4080

```

gtatacatgt atgttttgaa ggcaaggcaa aaaagattag aggatgaaga gatgaagtgg 4140  
 ggcacccctg aacttctatt ctctcaaaca tagtcatctt ccccatgtc ctcaggtgtg 4200  
 acaggatacc tgaaaattga tagcagtggc gatcgggaaa cagacttctc cctctgggat 4260  
 atggatcccc agaatggtgc cttcagggtg agtttgtgca cccagaagac agtgccaatt 4320  
 ccaaatgaca tctcacctc ctacttcccc cccacagccc tgccaggga cctgtttatc 4380  
 ctgtagccat tccaccatgc ctggacactt acaagagccc tggataaaac agaccagct 4440  
 ccagtctggg gaagccacca gaatgatagg gactcacagg catcacactt ggggagcccc 4500  
 atgcctgagg agggagcaca agcctgccct cggggagctc cgaagggagg caggcaggac 4560  
 cgctccctag cagagacagg gctgtgaaag atgcacatta cacagctctg caagcgagca 4620  
 gggacaggaa ggcgctgagg ccaatggcca caagggacag gtcattcaga gaaggcctcc 4680  
 tggaagacgg gcacatggac tgggcctgcg aatgtaggct aaggtgaaca ttaccttctc 4740  
 ctgttttcta ccaagaaaat aagtagagaa aaatcaatgc ttggttggtg cttcaaccaa 4800  
 gattataaac tccctgagtg tagagatcgg gttctaaatg gagttttctt tataaaccct 4860  
 ttgatagttt tcaggtgttt ccacttgagt actatgtgtg tggatgagg tctgtgtcc 4920  
 agttgcagtg gggacttggt aagcagggtg caaccagat atatatgtag gctctagaag 4980  
 cagagctggg gtaggtggga ggtgagactg ctgcactcac agcatgcctt ccccgaggc 5040  
 cctggcctag ccaccactcc tgcctccct taggttgtac tgaactaaa tgggacttcc 5100  
 caagagctgg tggctgtgtc ggggcgcaaa ctgaactggc ccctggggtg ccctcctcct 5160  
 gacatcccca aatgtggctt tgacaacgaa gaccagcat gcaaccaagg tgactgcccc 5220  
 ttgccttcca ggcctccatc ccagagatgc tgcactctc ccctaagcac agtcgagtag 5280  
 gtgctcctgt cccatgtctg gggctttctg gagaatgact cctgcctttt tcttcccttc 5340  
 atccatcatc ccagttcact gatggactat tagaaagttc ttcctcctgc tgtctaacc 5400  
 aaatctctct tgcctgaata tggactctct cctgcagatc acctttccac cctggaggtg 5460  
 ctggcttttg tgggcagcct ctccctgtct ggcatctga ttgtctcctt cttcatatac 5520  
 aggtgagctg tgatgtgggg ggttgagtga ggctggggga cccggagaac caagagcaga 5580  
 ggaggcgggtg gggaccacga gggagaggg caggggtgaa ggggcagcag gggaaaacca 5640  
 agggagatga ggaagaaagg aggcctaaaa gccagaggag aaagaaagag aagggaatgg 5700  
 cagggcgagg ggaggagaca aggataggaa tggccaagga gagtcagaaa gatccaagaa 5760  
 gcagagaagt tgatgggtga catcataggg gcgtggactg gttttccttg ctactcttgc 5820  
 aggccagata ggaagcaact ttctgaacct ttgcaatcat gccatgtta gctgaggagg 5880  
 gtgagccctg gtgtgtgcca ggtgcccaac ctagaatgga gaaggagct gaatgagcct 5940  
 tgttccctgc gtccagtggg ggctaaaatg aagtacagga ggagttaatg atatacaaaa 6000  
 gcaaggaggg aggggagaaa aatcactgct ggttgagcat ataatgtgtg ccaggcactt 6060  
 ccacgtacac tatttctttc tttctttttt tttttttttt tttttttttg agacggagtc 6120  
 tcgctctgtt gccagactgg agtgacagtg catgatctag gctcactgca acctccgct 6180  
 ccagtttca agcaattctc ctgcctcagc ctcccatgta gctgggacta caggcacatg 6240  
 ccaccagct cagctaattt ttgtattttt agtagagaca gggtttcacc atgttgcca 6300  
 ggatggtctc gatctcttga cctcatgatc caccacctt ggctcccaa agtgctggga 6360  
 ttacaggcat gagccactgt gcctggcctc atgttacta tttcttttca ttcttataat 6420  
 agttaagaat gaaatagata ttgcggcctc attcccaagt aaggacattg aggtgattcc 6480  
 cccaaggtcc ccagtaaggc agaatttccc ccagccatcc tgattctcag tccagaggat 6540  
 agaattcccc ctccatctct gagtgcattg tgtggtccca cggtctctgag gaggggctgc 6600  
 tgagcacct gccctgggtc agcggctcag ccacaggctc agatgcagcc ttcgatatccc 6660  
 aggaagatgc agctggagaa ggaactggcc tcggagctgt ggcgggtgcg ctgggaggac 6720  
 gttgagccca gtagccttga gaggcacctg cggagtgcag gcagccggct gacctgagc 6780  
 ggggtaagaa cgctggtgtt tgtgttgggg ggcaataaag gagaggtggg tacaaggggc 6840  
 agtgctctgag ggataggtaa gggagcagga ttctagtccc agctctgctt tcaattgctg 6900  
 tgtgacctg agcgactcat agtccctctc cgagactgtc tcagatgatg attacagcag 6960

cagagcctcc	ctcacagggc	totttttaaag	gtcagaggag	atagtacctg	tgaaaacact	7020
ttaaaaaaaa	aaaaagtaaa	tgaggaggaa	attttatgat	gtggaacata	aagcagggtg	7080
ggccaggcac	agtggctcac	atctgcaatc	ccagcacttt	gggagaccga	ggcaggagga	7140
ttgcttgtgc	ctgggagttc	aagaccagcc	tgggcaacag	agcaagacat	cgtctctaca	7200
aagaatacaa	agattagcag	ggcatggtgg	cgcatacctg	tagtcccagc	tactctggag	7260
gctgaggtga	aaggatcatc	tgagcccagg	agtctgaggc	ggcagtgacc	taggatagca	7320
ccactgcact	ccagcctgga	tgacacaatg	atactacatc	tcaaaaaaaaa	acccaacaac	7380
aaaaaggaag	ggtgacacaa	agataaggca	ggataaggca	gggaaataaa	gaccagagca	7440
caagcaatca	ggatgcagac	tgggcccacc	ggctgaccat	tcctcctgct	ctccctcctt	7500
tcagagaggc	tccaattacg	gtccctgct	aaccacagag	ggccagttcc	aagtctttgc	7560
caagacagca	tattataagg	tgggcctggg	gaaagatcac	tgggccttgg	gactggggca	7620
ggagtgtact	ctgatggagg	actggtgggg	ggttctgagg	gaaggagtaa	gctggtgggg	7680
agcagcagat	gggggccctg	ggggtgggct	attgggaaca	agtgagggtc	ctgagggcag	7740
ggatgggctg	tcgggagcag	ctggaattcc	caggacatgg	gaccatgctc	ttcacagtga	7800
cagtctccat	tccatgccc	gggcaacctc	gtggctgtga	aacgtgtgaa	ccgtaaacgc	7860
attgagctga	cacgaaaagt	cctgtttgaa	ctgaagcatg	taatgtgggg	agtgaggcag	7920
tggcatggag	aaggggccct	cggggacgca	agggagactg	gccaacagaa	ctagttatgg	7980
agggacctca	gggtacccc	agaaaggggc	agggactgga	gccctggatg	accttcatct	8040
tgtggtggag	tgggggtatc	ctaagtagga	gaagagacca	ctgagataac	ctggagggaat	8100
cttgaggggc	catatgtgat	gtccctgggg	gagagagggc	ttaggatgcc	agagggagta	8160
ggagcagatt	ctggggaggg	tgggctaaag	gacatgggtg	ggaatcacca	gggaagatct	8220
tagtgatggt	tgcagaaagt	gaataaggag	ttaagaagag	tgagggtccc	tgaagctagt	8280
gagcagcttg	gtgaggagcg	aggtctctgt	caagctcctg	atgctgggtc	cacttgcaga	8340
tgcgggatgt	gcagaatgaa	cacctgacca	ggtttgtggg	agcctgcacc	gaccccccca	8400
atatctgcat	cctcacagag	tactgtcccc	gtgggagcct	gcagggtagg	gggacaaggg	8460
gtgtcaagaa	acctgggttc	tagccctggc	tctgccctg	actggccata	agaccccagg	8520
catgcctcgc	cctctttctg	acctttctgg	ccccatctgt	aaaaatggga	gttggggaag	8580
ggcagtggca	ctagagtcaa	tccaaagttt	tgtcctgttc	taccagttca	catcagtagg	8640
accttgcacc	ctcctccaac	tcacaggggg	atctgcaggg	gattggtctt	gactcttatt	8700
gccccagcag	gacattctgg	agaatgagag	catcacctg	gactggatgt	tcgggtactc	8760
actaccaat	gacatcgtea	aggtatgcc	ctaagcacct	attggatgtg	tagagcaggg	8820
gccaggcatg	cttctcctgg	ccacgggtgt	aggtcccact	cctggccaat	acctctgccc	8880
actcacattt	ccagggcattg	ctgtttctac	acaatggggc	tatctgttcc	catgggaacc	8940
tcaagtcato	caactgcgtg	gtagatgggc	gctttgtgct	caagatcacc	gactatgggc	9000
tggagagctt	cagggacctg	gacccagagc	aaggacacac	cgtttatgcc	agtgagcctt	9060
gactcttgaa	cctaacacct	gccccagca	ccaccagta	gggagactga	tgcaaggcct	9120
ctgatgggct	tgggcatgct	tgtcctgact	ccagcctcaa	ttcattcacc	catgaaaaag	9180
ggaaggccag	acgaagtggg	ttctaaggcc	tcctctagct	ctaacactct	gtgatgcato	9240
cagatcagtt	tcggccacac	ccttgtttcc	ccctcacccc	ttagctttgg	gtccctcac	9300
tcggtgacta	ccgacctctg	acccacagaa	aagctgtgga	cggccccctga	gtccctgcga	9360
atggcttcac	ccctgtgcg	gggtccccag	gctggtgacg	tatacagctt	tgggatcatc	9420
cttcaggaga	ttgccctgag	gagtggggtc	ttccacgtgg	aaggtttgga	cctgagcccc	9480
aaaggtgaga	ggagcacacc	ttccttaaac	ccagccacag	tctcaacgaa	ccccagcccc	9540
agggagaggg	tcccctggca	gcaccaccac	accttccttc	tgtaatgggg	ttcagtcacc	9600
accttttgac	ccattgctgc	cagtgaccag	tccccgcgcc	ccatgccttg	gtcttggaact	9660
tcccctgcc	tctcagctgg	ttgccccagt	ctctcactag	gcccttggcc	agccccaccc	9720
ctcagctcct	ctacccccca	atacagagat	catcgagcgg	gtgactcggg	gtgagcagcc	9780
ccccttcggg	ccctccctgg	ccctgcagag	tcacctggag	gagttggggc	tgctcatgca	9840

```

gcggtgctgg gctgaggacc cacaggagag gccaccattc cagcagatcc gcctgacgtt 9900
gcgcaaattt aacagggtccc tgggtgtttgt catggatccc ccaggccctt cctccacagc 9960
caccattttac ctaatgcttc tggctctggc ttatcccagc agtggcagag ggagaccact 10020
cacctcctcc ctgtacatag tcagctccag ctcagcacag cctcatgacc ctcttcgcaa 10080
gtacagcatg actcagctgt cccacacagtc ccctgccatt catgcccctt cctccacca 10140
tcgacacccc acacccttcc tgcccactcg ccttgctggc ctctagactt ctcagcagtg 10200
tgtaggatag atgggcctcc cgccctcctgc cctgtaggct cttggccctc caggggagct 10260
cctgccccac cccttgattt cccttcccca gcgtgccac caggcccagt tcctccagac 10320
acacccttct gtggacatca ctttgtccgc aattgaccct tgtcattctc cacctccttt 10380
acctccttct aactcactgg gttcaacaaa gatgaacaaa atgtccatat gtctgaagct 10440
tcatacttga ccttggggtc tcagaaaaga attgaacttt cttccttctg ttttcccctg 10500
ctccccggta tcctgctatg ccctcaacc tcagcgtctc tagagacctc actgcagtct 10560
ggagggggaa gtgcctaggg gcgggcgctc acgtaggctg tgctgctcct ctcttaccac 10620
ccccaccgcc accctctgcc ccaggggaga acagcagcaa catcctggac aacctgctgt 10680
cccgcatgga gcagtacgcg aacaatctgg aggaactggt ggaggagcgg acccaggcat 10740
acctggagga gaagcgcaag gctgaggccc tgctctacca gatcctgcct cagtgcgtgc 10800
ctgagtctgg ggaccccccc caacacaaag cccctgtccc gacccccaac tctgacctg 10860
cacctgcctt gacccttag ctcagtggct gagcagctga agcgtgggga gacggtgcag 10920
gccgaagcct ttgacagtgt taccatctac ttcagtga ca ttgtgggtt cacagcgtg 10980
tcggcggaga gcacacccat gcaggtaggc cagggttcag ccacagggtc caggcaagct 11040
cagcatctgg atcccaccag acctgccttc tggttctgct ttaccacct gacccaggt 11100
ggggtccctt acttctctgc tctcttagct tctcttccct tcagggtggt gacctgctc 11160
aatgacctgt aacttgctt tgatgctgtc atagacaact ttgatgtgta caaggtgagg 11220
gtgggagtgg ggatgggaag ggacagacag acatggacaa ggtcagaaaa agatgagggg 11280
taggcagaat gatgtggagt cttaagagag gagatcgggg acacgggcag agacagtgc 11340
acagggagac ccgggaacag gcagagaacc catgtgggat gggggatgag caaagacaga 11400
tgagggtaca gaatgacaga cgctgcacc ggtgtgacgg tgtggccggc cgcacagtgc 11460
cagccgtcaa gtccctgcacc ccctcgccac tcccacaggt ggagacaatt ggcgatgcct 11520
acatggtggt gtcagggtc cctgtgcgga acgggcggct acacgcctgc gaggtagccc 11580
gcatggcctt ggcactgctg gatgctgtgc gtccttccg aatccgccac cggccccagg 11640
agcagctgcg cttgcgcatt ggcattccca caggtaaggc cactgaaggt gcaggcgggc 11700
atccagaggc caaggctttg caagggaac ttgtccctg gccagcccc tcgccctttc 11760
atctctctct ctctctctct ctctctctct ctctctctct gtctctctct ctctctctct 11820
ctctctctct ctcacacaca cacacacaca cacacacaga gctgggacct cagatcctgc 11880
ctctgctctg tcttggattg tccacctacc tcccttaaca cccctccctc cctcactgc 11940
tgatgggtc tgctccttcc cttgctcctc ccaggacctg tgtgtgctgg agtgggtggga 12000
ctgaagatgc ccgttactg tctctttggg gatacagtca acacagcctc aagaatggag 12060
tctaattggg aaggtacagt gcccctcct agagggaatg gggagggcag ggtggctgag 12120
ggaaatgcca tcctggggca gcctgtgcct gcacagcccg tttcagctcc tagccctttc 12180
gcctcccaag tcccccttct cataatatta agagttcaac ctgggctcat caacttgact 12240
gtaaccagag actcaggttc ctgctgcccc tcttgtcaaa cgatgtaaaa gtatttccgg 12300
gccagtgcgt gagagttccc agcaggaatc tgattttaag accctctgtg ggccggcgct 12360
ggtgactcac acctgtgatc ccagcacttt gggaagctga ggcaggcgga tcacctgagg 12420
tcgggggttt cgagaccagc ctgaccaaca tgatgaaatc ccgtctctac taaaaatata 12480
aaaaactagc caggtgtgat ggcaggctcc tgtaatccca gctacttggg aggcttgagg 12540
cagaagaatt gcttgaaccc gggaggcaga ggttgcgatg agccaagatt acaccacgca 12600
ccccagcttg ggcaataaga gttaaactct gtctcaaaaa aaaaaaaaaa aaaaaaaaaa 12660
agggcctct gctccacctt tgatgtggta aagatggctt cagagccagc ataagtgagg 12720

```

ctgtgaatct	cagctccaca	gctggctgtg	tgctagtttg	ctatacctct	ctgagccatg	12780
gttttctca	tctgtaaaaa	gagggaaaaa	atctatctca	caggaattat	gtgagaaacc	12840
cattaaaaat	gtctaccaca	taattgtcat	ttaaacttttc	caagccttag	cggattatct	12900
gtaaaatgat	gtctatctca	ggattgcaag	aagcctagca	caaaccctgg	taccagcag	12960
gcacctata	aattcttact	cctacccgcc	ccttgctctt	gcctcctggt	tatcttctat	13020
ccttctgctg	tattcgacac	aattcaatgc	agtaaacatt	tattgagtga	ctactgagt	13080
ccaggccctg	ggatagtaac	atggcccaga	tccagagtta	gctgagaaat	tcatgtggac	13140
cccatctaaa	ccttatgggtg	aaagaaaggc	tgcttgggag	ccagtcctgg	gagccagag	13200
ggatctagtt	cggcaaatat	tccctgggca	ctatttgggg	gctgcagagt	cagcccttgt	13260
tgagggtcca	gtcctcaagg	agcacattcc	cagaaatggt	cacattctgg	cgtgggggtg	13320
ctgtaatccc	agcacttttg	gaggccgagg	tgggcagatc	acttgaggcc	aggagtggag	13380
actagcctgg	ccaacatggt	gacctcctgt	ctctactaaa	aatacaaaaa	attagetggg	13440
cgtggtggca	cgtgcccgta	atcccagcta	ctcaggaggc	ttgagacatg	aaaatcactt	13500
gaaccagga	ggtggatggt	gcagtgaacc	gagactgcac	ccctgggcaa	cagagcgaga	13560
ctctgtctca	aaaaaaaaaa	agagagaaag	aaagaaaaga	aaagaaagaa	actgttaaac	13620
acaacaaggc	cactgtgatt	gatgcaaacc	ccagaagtag	ggacatgagt	tcagacagt	13680
gtcaaagaga	gggtgtggca	atattgggcc	ccactccatc	actgacctcc	tcagccactt	13740
gggcagatca	ccctgggcct	cagttcctcg	gccacaaaat	gaggggtatag	catgaaatca	13800
tgaaagcaac	aattttacata	gtgcttccta	ggtagcacat	tccgtttgaa	tacttttatgg	13860
atgttaaatt	taatcctcac	aacaagggtt	tgagatgggt	actgacacta	tcagcatttt	13920
acagattagg	aaaatgaagc	agagagaatt	tattttacat	acctaagcaa	gtatccaagc	13980
tgaggttcat	actgaggcag	tgcaggatcc	aaagtgccag	ctcctaacca	ccatgctgtg	14040
tagagccggg	tgacactcca	gagagtgtcg	tccaacagga	tgttccatag	tcagtgaat	14100
gttctgtatt	ctgtgctgtc	caatacagta	gcctctaggc	acatatggct	acttatcact	14160
ggaaatgtga	cgggtgcaac	tgaggccctg	atTTTTTTTT	TTTTTTTgga	gacagagttt	14220
cgtctgtcg	cccagcctgg	atggagtgtg	gtgggtgcaat	ctcggctcac	tgcaacctcc	14280
gcctcccagg	ttcaagcgat	tctcctgcct	cagcctccca	agtagctgga	attacagggt	14340
agtgccacca	cacacagcta	atTTTTgtat	TTTTtagtaga	gacgggggtt	cgccatattg	14400
gccaggatgg	tctcgaactc	ctggcctcaa	gtgatcctcc	tgcctcagcc	tcccaaagt	14460
ctgggattac	aggtgtgagc	cacagcacc	agcctgaatt	tttaactgta	tttagtttaa	14520
attaatttaa	gttgaaacag	gcacatgtga	ttagtggcta	ctgtattgga	ttacacagct	14580
ccagagttct	aaatgagagg	ctaattgtgt	cacgcactac	attcaggggg	tggggccct	14640
ctgagctaga	gggcttctcg	gccccaaaaga	gggagagagg	gtacctgtcc	acctgtccac	14700
ccccacagtc	cctggtctct	tttgctctca	cttctctgct	ctcctctctc	acattgctca	14760
ccttcccttc	tccctgtcc	taccagccc	tgaagatcca	cttgtcttct	gagaccaagg	14820
ctgtcctgga	ggagtttgg	ggtttcgagc	tggagcttcg	aggggatgta	gaaatgaagg	14880
tagagcgaga	agcctctgcc	ctccccacct	tttgggggtcc	tagagggagt	tacccttctc	14940
aagcagccga	tgccactccc	atccctaagg	ctctcatctg	actggggaaa	gggcatgtgc	15000
cactccccag	cccatcctct	tttttccctc	cagggcaaag	gcaagggttcg	gacctactgg	15060
ctccttgggg	agagggggag	tagcaccgga	ggctgacctg	cctcctctcc	tatccctcca	15120
cacctccct	accctgtgcc	agaagcaaca	gaggtgccag	gcctcagcct	caccacagc	15180
agccccatcg	ccaaaggatg	gaagtaattt	gaatagctca	ggtgtgctga	ccccagtga	15240
gacaccagat	aggacctctg	agagggggact	ggcatggggg	gatctcagag	cttacaggct	15300
gagccaagcc	cacggccatg	cacagggaca	ctcacacagg	cacacgcacc	tgctctccac	15360
ctggactcag	gccgggctgg	gctgtggatt	cctgateccc	tccctcccc	atgctctcct	15420
ccctcagcct	tgctacctcg	tgacttactg	ggaggagaaa	gagtcacctg	aaggggaaca	15480
tgaaaagaga	ctaggtgaag	agagggcagg	ggagcccaca	tctggggctg	gcccacaata	15540
cctgctcccc	cgacccccctc	caccagcag	tagacacagt	gcacagggga	gaagaggggt	15600

```

jgcgcagaag ggttgggggc ctgtatgcct tgcttctacc atgagcagag acaattaaaa 15660
:ctttattcc agtgacagtg tctcttcttg agggagagag ggttgccaga aaacagtcag 15720
:tctccactc tctacttcaa ataagactca cttcttgttc tacaagggtc tagaaggaaa 15780
agtaaaaaaa aaagactctc gattcttaac 15810

```

&lt;210&gt; 2

&lt;211&gt; 3186

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

atgcgcggggc cccggcgccc cgetggctcc cgctgcgcgc tgctcctgct cctgctgctg 60
ccgccgctgc tgctgctgct ccggggcagc cagcggggca acctgacggg agccgtggta 120
ctgccgctgg ccaatacctc gtacccttg tgctgggcgc gcgtgggacc cgccgtggag 180
ctggcccttg cccaggtgaa ggcgcgcccc gacttgctgc cgggctggac ggtccgcacg 240
gtgctgggca gcagcgaaaa cgcgctgggc gtctgctccg acaccgcagc gccctggcc 300
jcggtggacc tcaagtggga gcacaacccc gctgtgttcc tgggccccgg ctgcgtgtac 360
jccgccgccc cagtggggcg cttcacgcgc cactggcggg tcccgtgct gaccgccggc 420
jccccggcgc tgggcttcgg tgtcaaggac gagtatgcgc tgaccacccg cgcggggccc 480
agctacgcca agctggggga cttcgtggcg gcgctgcacc gacggctggg ctgggagcgc 540
caagcgctca tgctctacgc ctaccggccg ggtgacgaag agcactgctt cttcctcgtg 600
jaggggctgt tcatgcggtt ccgcgaccgc ctcaatatta cgggtggacca cctggagttc 660
jccgaggacg acctcagcca ctacaccagg ctgctgcgga ccatgccgcg caaaggccga 720
jttatctaca tctgcagctc ccctgatgcc ttcagaaccc tcatgctcct ggccctggaa 780
jctggcttgt gtggggagga ctacgttttc ttccacctgg atatctttgg gcaaagcctg 840
caaggtggac agggccctgc tccccgcagg ccctgggaga gaggggatgg gcaggatgtc 900
agtccccgc aggcctttca ggctgccaaa atcattacat ataaagaccc agataatccc 960
jagtacttgg aattcctgaa gcagttaaaa cacctggcct atgagcagtt caacttcacc 1020
atggaggatg gcctggtgaa caccatccca gcatccttcc acgacgggct cctgctctat 1080
atccaggcag tgacggagac tctggcacat gggggaactg ttactgatgg ggagaacatc 1140
actcagcgga tgtggaaccg aagctttcaa ggtgtgacag gatacctgaa aattgatagc 1200
agtggcgatc gggaaacaga cttctccctc tgggatatgg atcccgagaa tggcgccctc 1260
agggttgtac tgaactacaa tgggacttcc caagagctgg tggctgtgtc ggggcgcaaa 1320
ctgaactggc ccctggggta ccctcctcct gacatcccca aatgtggctt tgacaacgaa 1380
gaccagcat gcaaccaaga tcacctttcc accctggagg tgctggcttt ggtgggcagc 1440
ctctccttgc tcggcattct gattgtctcc ttcttcatat acaggaagat gcagctggag 1500
aaggaactgg cctcggagct gtggcgggtg cgctgggagg acgttgagcc cagtagcctt 1560
gagaggcacc tgcggagtgc aggcagccgg ctgacctga gcgggagagg ctccaattac 1620
jgctccctgc taaccacaga gggccagttc caagtctttg ccaagacagc atattataag 1680
jgcaacctcg tggctgtgaa acgtgtgaac cgtaaacgca ttgagctgac acgaaaagtc 1740
ctgtttgaac tgaagcatat gcgggatgtg cagaatgaac acctgaccag gtttgtggga 1800
jctgcaccg acccccccaa tatctgcac ctcacagagt actgtccccg tgggagcctg 1860
caggacattc tggagaatga gagcatcacc ctggactgga tgttcgggta ctactcacc 1920
aatgacatcg tcaagggcag gctgtttcta cacaatgggg ctatctgttc ccatgggaac 1980
ctcaagtcac ccaactgcgt ggtagatggg cgctttgtgc tcaagatcac cgactatggg 2040
ctggagagct tcagggacct ggaccagag caaggacaca ccgtttatgc caaaaagctg 2100
tggacggccc ctgagctcct gcgaatggct tcacccctg tgcggggctc ccaggctggt 2160

```

```

gacgtataca gctttgggat catccttcag gagattgcc tgaggagtgg ggtcttccac 2220
gtggaaggtt tggacctgag ccccaaagag atcatcgagc gggtgactcg gggtagcag 2280
cccccttcc ggccctccct ggccctgcag agtcacctgg aggagttggg gctgctcatg 2340
cagcgggtgct gggctgagga cccacaggag aggccacat tccagcagat ccgcctgacg 2400
ttgcgcaaat ttaacaggga gaacagcagc aacatcctgg acaacctgct gtcccgcag 2460
gagcagtacg cgaacaatct ggaggaactg gtggaggagc ggaccaggc atacctggag 2520
gagaagcgca aggetgaggc cctgctctac cagatcctgc ctactcagt ggctgagcag 2580
ctgaagcgtg gggagacggt gcaggccgaa gcctttgaca gtggttaccat ctacttcagt 2640
gacattgtgg gtttcacagc gctgtcggcg gagagcacac ccatgcaggt ggtgacctg 2700
ctcaatgacc tgtacacttg ctttgatgct gtcatagaca actttgatgt gtacaagggtg 2760
gagacaattg gcgatgccta catggtggtg tcagggctcc ctgtgcggaa cgggcggcta 2820
cacgcctgcg aggtagcccg catggccctg gcactgctgg atgctgtgcg ctcttccga 2880
atccgccacc ggccccagga gcagctgcgc ttgcgcattg gcattccacac aggacctgtg 2940
tgtgctggag tgggtgggact gaagatgcc cgttactgtc tctttgggga tacagtcaac 3000
acagcctcaa gaatggagtc taatggggaa gccctgaaga tccactgtc ttctgagacc 3060
aaggctgtcc tggaggagtt tgggtggttc gagctggagc ttcgagggga tgtagaaatg 3120
aaggggcaaag gcaagggttcg gacctactgg ctcttgggg agagggggag tagcaccgca 3180
ggctga 3186

```

&lt;210&gt; 3

&lt;211&gt; 1061

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

```

Met Pro Gly Pro Arg Arg Pro Ala Gly Ser Arg Leu Arg Leu Leu Leu
  1              5              10              15

```

```

Leu Leu Leu Leu Pro Pro Leu Leu Leu Leu Leu Arg Gly Ser His Ala
      20              25              30

```

```

Gly Asn Leu Thr Val Ala Val Val Leu Pro Leu Ala Asn Thr Ser Tyr
      35              40              45

```

```

Pro Trp Ser Trp Ala Arg Val Gly Pro Ala Val Glu Leu Ala Leu Ala
      50              55              60

```

```

Gln Val Lys Ala Arg Pro Asp Leu Leu Pro Gly Trp Thr Val Arg Thr
      65              70              75              80

```

```

Val Leu Gly Ser Ser Glu Asn Ala Leu Gly Val Cys Ser Asp Thr Ala
      85              90              95

```

```

Ala Pro Leu Ala Ala Val Asp Leu Lys Trp Glu His Asn Pro Ala Val
      100             105             110

```

```

Phe Leu Gly Pro Gly Cys Val Tyr Ala Ala Ala Pro Val Gly Arg Phe

```

9

370	375	380
Trp Asn Arg Ser Phe Gln Gly Val Thr Gly Tyr Leu Lys Ile Asp Ser		
385	390	395 400
Ser Gly Asp Arg Glu Thr Asp Phe Ser Leu Trp Asp Met Asp Pro Glu		
405	410	415
Asn Gly Ala Phe Arg Val Val Leu Asn Tyr Asn Gly Thr Ser Gln Glu		
420	425	430
Leu Val Ala Val Ser Gly Arg Lys Leu Asn Trp Pro Leu Gly Tyr Pro		
435	440	445
Pro Pro Asp Ile Pro Lys Cys Gly Phe Asp Asn Glu Asp Pro Ala Cys		
450	455	460
Asn Gln Asp His Leu Ser Thr Leu Glu Val Leu Ala Leu Val Gly Ser		
465	470	475 480
Leu Ser Leu Leu Gly Ile Leu Ile Val Ser Phe Phe Ile Tyr Arg Lys		
485	490	495
Met Gln Leu Glu Lys Glu Leu Ala Ser Glu Leu Trp Arg Val Arg Trp		
500	505	510
Glu Asp Val Glu Pro Ser Ser Leu Glu Arg His Leu Arg Ser Ala Gly		
515	520	525
Ser Arg Leu Thr Leu Ser Gly Arg Gly Ser Asn Tyr Gly Ser Leu Leu		
530	535	540
Thr Thr Glu Gly Gln Phe Gln Val Phe Ala Lys Thr Ala Tyr Tyr Lys		
545	550	555 560
Gly Asn Leu Val Ala Val Lys Arg Val Asn Arg Lys Arg Ile Glu Leu		
565	570	575
Thr Arg Lys Val Leu Phe Glu Leu Lys His Met Arg Asp Val Gln Asn		
580	585	590
Glu His Leu Thr Arg Phe Val Gly Ala Cys Thr Asp Pro Pro Asn Ile		
595	600	605
Cys Ile Leu Thr Glu Tyr Cys Pro Arg Gly Ser Leu Gln Asp Ile Leu		
610	615	620
Glu Asn Glu Ser Ile Thr Leu Asp Trp Met Phe Arg Tyr Ser Leu Thr		

625	630	635	640
Asn Asp Ile Val Lys Gly Met Leu Phe Leu His Asn Gly Ala Ile Cys			
	645	650	655
Ser His Gly Asn Leu Lys Ser Ser Asn Cys Val Val Asp Gly Arg Phe			
	660	665	670
Val Leu Lys Ile Thr Asp Tyr Gly Leu Glu Ser Phe Arg Asp Leu Asp			
	675	680	685
Pro Glu Gln Gly His Thr Val Tyr Ala Lys Lys Leu Trp Thr Ala Pro			
	690	695	700
Glu Leu Leu Arg Met Ala Ser Pro Pro Val Arg Gly Ser Gln Ala Gly			
705	710	715	720
Asp Val Tyr Ser Phe Gly Ile Ile Leu Gln Glu Ile Ala Leu Arg Ser			
	725	730	735
Gly Val Phe His Val Glu Gly Leu Asp Leu Ser Pro Lys Glu Ile Ile			
	740	745	750
Glu Arg Val Thr Arg Gly Glu Gln Pro Pro Phe Arg Pro Ser Leu Ala			
	755	760	765
Leu Gln Ser His Leu Glu Glu Leu Gly Leu Leu Met Gln Arg Cys Trp			
	770	775	780
Ala Glu Asp Pro Gln Glu Arg Pro Pro Phe Gln Gln Ile Arg Leu Thr			
785	790	795	800
Leu Arg Lys Phe Asn Arg Glu Asn Ser Ser Asn Ile Leu Asp Asn Leu			
	805	810	815
Leu Ser Arg Met Glu Gln Tyr Ala Asn Asn Leu Glu Glu Leu Val Glu			
	820	825	830
Glu Arg Thr Gln Ala Tyr Leu Glu Glu Lys Arg Lys Ala Glu Ala Leu			
	835	840	845
Leu Tyr Gln Ile Leu Pro His Ser Val Ala Glu Gln Leu Lys Arg Gly			
	850	855	860
Glu Thr Val Gln Ala Glu Ala Phe Asp Ser Val Thr Ile Tyr Phe Ser			
865	870	875	880
Asp Ile Val Gly Phe Thr Ala Leu Ser Ala Glu Ser Thr Pro Met Gln			

[illegible]

```
<210> 4
<211> 15
<212> DNA
<213> Homo sapiens
```

<400> 4  
gatgcctr gg accgg

15

$\langle 210 \rangle$	5
$\langle 211 \rangle$	15
$\langle 212 \rangle$	DNA

<213> Homo sapiens

<400> 5

atgcctgsga ccggc

15

<210> 6

<211> 15

<212> DNA

<213> Homo sapiens

<400> 6

gccatgcygg ggccc

15

<210> 7

<211> 15

<212> DNA

<213> Homo sapiens

<400> 7

gccccggmgc cccgc

15

<210> 8

<211> 15

<212> DNA

<213> Homo sapiens

<400> 8

ccccggcset gggct

15

<210> 9

<211> 15

<212> DNA

<213> Homo sapiens

<400> 9

cgccaagygc tcatg

15

<210> 10

<211> 15

<212> DNA

<213> Homo sapiens

<400> 10

ctgactcycc gtctt

15

<210> 11

<211> 15

<212> DNA

<213> Homo sapiens

<400> 11

cctcagcrtc tgaaa

15

<210> 12

<211> 15

<212> DNA

<213> Homo sapiens

<400> 12

tcaccatsga ggatg

15

<210> 13

<211> 15

<212> DNA

<213> Homo sapiens

<400> 13

ctcctacwtc ccccc

15

<210> 14

<211> 15

<212> DNA

<213> Homo sapiens

<400> 14

gctcctgycc catgc

15

<210> 15

<211> 15

<212> DNA

<213> Homo sapiens

<400> 15

gtgatgtrgg gggtt

15

<210> 16  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 16  
acttgctrtg tgacc 15

<210> 17  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 17  
ataaggcwgg ataag 15

<210> 18  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 18  
tcggggaygc aaggg 15

<210> 19  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 19  
gactaccrac ctctg 15

<210> 20  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 20  
cattgctrcc agtga 15

<210> 21  
<211> 15  
<212> DNA

<213> Homo sapiens

<400> 21

ccatctcwgcc tgggtt

15

<210> 22

<211> 15

<212> DNA

<213> Homo sapiens

<400> 22

cgttgcgyaa attta

15

<210> 23

<211> 15

<212> DNA

<213> Homo sapiens

<400> 23

cagcagtrgc agagg

15

<210> 24

<211> 15

<212> DNA

<213> Homo sapiens

<400> 24

caccagayct gcctt

15

<210> 25

<211> 15

<212> DNA

<213> Homo sapiens

<400> 25

gcgcctgatg cctrg

15

<210> 26

<211> 15

<212> DNA

<213> Homo sapiens

<400> 26

cagcggccgg tccya

15

<210> 27

<211> 15

<212> DNA

<213> Homo sapiens

<400> 27

cgccctgatgc ctgsg

15

<210> 28

<211> 15

<212> DNA

<213> Homo sapiens

<400> 28

tcagcggccg gtcsc

15

<210> 29

<211> 15

<212> DNA

<213> Homo sapiens

<400> 29

gctgaggcca tgcyg

15

<210> 30

<211> 15

<212> DNA

<213> Homo sapiens

<400> 30

gcgccggggc cccrg

15

<210> 31

<211> 15

<212> DNA

<213> Homo sapiens

<400> 31

gccggggccc cggmg

15

<210> 32  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 32  
gagccagcgg ggckc 15

<210> 33  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 33  
ccggcgcccc ggcsc 15

<210> 34  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 34  
caccgaagcc cagsg 15

<210> 35  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 35  
tgggagcgcc aagyg 15

<210> 36  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 36  
gtagagcatg agcrc 15

<210> 37  
<211> 15  
<212> DNA

<213> Homo sapiens

<400> 37

ctctctctga ctcyc

15

<210> 38

<211> 15

<212> DNA

<213> Homo sapiens

<400> 38

tggagaaaga cggrg

15

<210> 39

<211> 15

<212> DNA

<213> Homo sapiens

<400> 39

tgtgtccctc agcrt

15

<210> 40

<211> 15

<212> DNA

<213> Homo sapiens

<400> 40

gaattctttc agayg

15

<210> 41

<211> 15

<212> DNA

<213> Homo sapiens

<400> 41

tcaacttcac catsg

15

<210> 42

<211> 15

<212> DNA

<213> Homo sapiens

<400> 42

ccaggccatc ctcsa

15

&lt;210&gt; 43

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 43

ctcacctcc tacwt

15

&lt;210&gt; 44

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 44

gctgtggggg ggawg

15

&lt;210&gt; 45

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 45

gtaggtgctc ctgyc

15

&lt;210&gt; 46

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 46

ccctcagcat gggrc

15

&lt;210&gt; 47

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 47

tgagctgtga tgtrg

15

<210> 48  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 48  
tcactcaacc cccya 15

<210> 49  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 49  
gctttcactt gctr 15

<210> 50  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 50  
gctcaaggtc acaya 15

<210> 51  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 51  
acaaagataa ggcwg 15

<210> 52  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 52  
ccctgcctta tccwg 15

<210> 53  
<211> 15  
<212> DNA

<213> Homo sapiens

<400> 53

gggccctcgg ggayg

15

<210> 54

<211> 15

<212> DNA

<213> Homo sapiens

<400> 54

cagtctccct tgcr

15

<210> 55

<211> 15

<212> DNA

<213> Homo sapiens

<400> 55

ctcggtgact accra

15

<210> 56

<211> 15

<212> DNA

<213> Homo sapiens

<400> 56

gtgggtcaga ggtyg

15

<210> 57

<211> 15

<212> DNA

<213> Homo sapiens

<400> 57

ttgaccatt gctrc

15

<210> 58

<211> 15

<212> DNA

<213> Homo sapiens

<400> 58

gactgggtcac tggga

15

<210> 59

<211> 15

<212> DNA

<213> Homo sapiens

<400> 59

cccctgccat ctcwg

15

<210> 60

<211> 15

<212> DNA

<213> Homo sapiens

<400> 60

tggggcaacc agcwg

15

<210> 61

<211> 15

<212> DNA

<213> Homo sapiens

<400> 61

gcctgacgtt gcgga

15

<210> 62

<211> 15

<212> DNA

<213> Homo sapiens

<400> 62

acctgttaaa tttrc

15

<210> 63

<211> 15

<212> DNA

<213> Homo sapiens

<400> 63

ttatcccagc agtrg

15

<210> 64  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 64  
ggctctccctc tgcya 15

<210> 65  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 65  
ggatcccacc agayc 15

<210> 66  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 66  
aaccagaagg cagrt 15

<210> 67  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 67  
cctgatgcct 10

<210> 68  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 68  
cggccggtcc 10

<210> 69  
<211> 10  
<212> DNA

<213> Homo sapiens

<400> 69

ctgatgcctg

10

<210> 70

<211> 10

<212> DNA

<213> Homo sapiens

<400> 70

gcggccggtc

10

<210> 71

<211> 10

<212> DNA

<213> Homo sapiens

<400> 71

gaggccatgc

10

<210> 72

<211> 10

<212> DNA

<213> Homo sapiens

<400> 72

ccggggcccc

10

<210> 73

<211> 10

<212> DNA

<213> Homo sapiens

<400> 73

ggggccccgg

10

<210> 74

<211> 10

<212> DNA

<213> Homo sapiens

<400> 74

ccagcggggc

10

&lt;210&gt; 75

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 75

gcgccccggc

10

&lt;210&gt; 76

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 76

cgaagcccag

10

&lt;210&gt; 77

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 77

gagcgccaag

10

&lt;210&gt; 78

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 78

gagcatgagc

10

&lt;210&gt; 79

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 79

tctctgactc

10

<210> 80  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 80  
agaaagacgg

10

<210> 81  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 81  
gtccctcagc

10

<210> 82  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 82  
ttctttcaga

10

<210> 83  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 83  
acttcaccat

10

<210> 84  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 84  
ggccatcctc

10

<210> 85  
<211> 10  
<212> DNA

<213> Homo sapiens

<400> 85

accctcctac

10

<210> 86

<211> 10

<212> DNA

<213> Homo sapiens

<400> 86

gtggggggga

10

<210> 87

<211> 10

<212> DNA

<213> Homo sapiens

<400> 87

ggtgctcctg

10

<210> 88

<211> 10

<212> DNA

<213> Homo sapiens

<400> 88

tcagcatggg

10

<210> 89

<211> 10

<212> DNA

<213> Homo sapiens

<400> 89

gctgtgatgt

10

<210> 90

<211> 10

<212> DNA

<213> Homo sapiens

<400> 90

ctcaaccccc

10

&lt;210&gt; 91

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 91

ttcacttgct

10

&lt;210&gt; 92

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 92

caaggtcaca

10

&lt;210&gt; 93

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 93

aagataaggc

10

&lt;210&gt; 94

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 94

tgccttatcc

10

&lt;210&gt; 95

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 95

ccctcgggga

10

<210> 96  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 96  
tctcccttgc 10

<210> 97  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 97  
ggtgactacc 10

<210> 98  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 98  
ggtcagaggt 10

<210> 99  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 99  
acccattgct 10

<210> 100  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 100  
tggtcactgg 10

<210> 101  
<211> 10  
<212> DNA

<213> Homo sapiens

<400> 101  
ctgccatctc

10

<210> 102  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 102  
ggcaaccagc

10

<210> 103  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 103  
tgacgttgcg

10

<210> 104  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 104  
tgttaaattt

10

<210> 105  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 105  
tcccagcagt

10

<210> 106  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 106

ctccctctgc

10

&lt;210&gt; 107

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 107

tcccaccaga

10

&lt;210&gt; 108

&lt;211&gt; 10

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 108

cagaaggcag

10

&lt;210&gt; 109

&lt;211&gt; 15810

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; (730)

&lt;223&gt; PS1: Polymorphic base G or A

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; (731)

&lt;223&gt; PS2: Polymorphic base G or C

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; (811)

&lt;223&gt; PS3: Polymorphic base C or T

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; (822)

&lt;223&gt; PS4: Polymorphic base C or A

&lt;220&gt;

&lt;221&gt; allele

<222> (1235)  
<223> PS5: Polymorphic base G or C

<220>  
<221> allele  
<222> (1351)  
<223> PS6: Polymorphic base C or T

<220>  
<221> allele  
<222> (2184)  
<223> PS7: Polymorphic base T or C

<220>  
<221> allele  
<222> (2472)  
<223> PS8: Polymorphic base A or G

<220>  
<221> allele  
<222> (2979)  
<223> PS9: Polymorphic base G or C

<220>  
<221> allele  
<222> (4345)  
<223> PS10: Polymorphic base T or A

<220>  
<221> allele  
<222> (5290)  
<223> PS11: Polymorphic base T or C

<220>  
<221> allele  
<222> (5537)  
<223> PS12: Polymorphic base G or A

<220>  
<221> allele  
<222> (6900)  
<223> PS13: Polymorphic base G or A

<220>  
<221> allele  
<222> (7410)  
<223> PS14: Polymorphic base A or T

<220>  
 <221> allele  
 <222> (7947)  
 <223> PS15: Polymorphic base C or T

<220>  
 <221> allele  
 <222> (9313)  
 <223> PS16: Polymorphic base G or A

<220>  
 <221> allele  
 <222> (9619)  
 <223> PS17: Polymorphic base G or A

<220>  
 <221> allele  
 <222> (9675)  
 <223> PS18: Polymorphic base A or T

<220>  
 <221> allele  
 <222> (9904)  
 <223> PS19: Polymorphic base C or T

<220>  
 <221> allele  
 <222> (10004)  
 <223> PS20: Polymorphic base G or A

<220>  
 <221> allele  
 <222> (11062)  
 <223> PS21: Polymorphic base C or T

<400> 109  
 ggatcccaaa ccagcacacc ttccctctt ccccgagga gaccaggtag gaggcgaggg 60  
 aaaaggtggg gcgcaagtgg gcccggttg ctccacaca caccctcgt tcagccgtcc 120  
 tttccatccc ggcgagggcg caccttcaga gggctctgtc ctccaaagag gtaggcgtgg 180  
 ggcggccgag accggggaag atggtccacg gggaagcgcg cgggctgggc ggcggggagg 240  
 aaggagtcta tgatcctgga ttggctctt tgtcactgag tctgggaggg gaagcggctg 300  
 ggagggaggg ttcggaactt ggctcgggtc ctccacggtt ccctcggat agccggagac 360  
 ttgggcccgc cggacgcccc ttctggcaca ctccctgggg caggcgctca cgcacgctac 420  
 aaacacacac tcctctttcc tcctcgcgc gccctctctc atccttcttc acgaagcgct 480  
 cactcgcacc ctttctctct ctctctctct cttaacacg cagcacact cccagttgtt 540  
 cacactcggg tcctctccag cccgacgttc tcctggcacc cacctgctcc gcggcgccct 600  
 gcacgcccc ctcggtcgcg ccccttgcg tctcgccca gaccgtcgca gctacagggg 660  
 gcctcgagcc cgggggtgag cgtccccgtc ccgtcctgc tccttcccat agggacgcgc 720

ctgatgcctr	sgaccggccg	ctgagcccaa	ggggaccgag	gaggccatgg	taggagcgct	780
cgcctgctgc	ggtgcccgt	gaggccatgc	yggggcccg	gmgcccgcgt	ggctcccgcc	840
tgcgctgct	cctgctcctg	ctgctgccgc	cgctgctgct	gctgctccgg	ggcagccacg	900
cgggcaacct	gacggtagcc	gtggtagctgc	cgctggccaa	tacctcgtac	ccctggctgt	960
gggcgcgcgt	gggaccgcc	gtggagctgg	ccctggccca	ggtgaaggcg	cgccccgact	1020
tgctgccggg	ctggacggtc	cgcacgggtgc	tgggcagcag	cgaaaacgcg	ctgggcgtct	1080
gctccgacac	cgcagcgccc	ctggcccgcg	tggacctcaa	gtgggagcac	aaccccgctg	1140
tgttcctggg	ccccggctgc	gtgtacgccg	ccgccccagt	ggggcgcttc	accgcgcact	1200
ggcgggtccc	gctgctgacc	gccggcgccc	cggcsctggg	cttcgggtgtc	aaggacgagt	1260
atgcgctgac	cacccgcgcg	gggcccagct	acgccaagct	gggggacttc	gtggcggcgc	1320
tgcaccgacg	gctgggctgg	gagcgccaa	ygtcatgct	ctacgcctac	cggcgggtg	1380
acgaagagca	ctgcttcttc	ctcgtggagg	ggctgttcat	gcgggtccgc	gaccgcctca	1440
atattacggt	ggaccacctg	gagttcgccg	aggacgacct	cagccactac	accaggctgc	1500
tgcggaccat	gccgcgcaaa	ggccgagggtg	agacgctggc	acaccccgctc	ccgcgcgtta	1560
gccgcagggc	ctccccctctg	acctgccgga	ggcatcgga	ctttctctct	catctggggg	1620
cactcttctt	tctcctcgcc	gttcttctt	ctactttcag	ctccctggcc	ctttctacag	1680
ctgagtttct	atttccctct	cttcttccgc	cacccccacc	acgtctctat	cctctcatct	1740
ccccgacccc	cactcattcc	ctccccacct	agcacagctc	ggttccggtc	cttttttccc	1800
tcccacattt	tctctcttcc	ctatagcctt	ctcccttctt	tcatcctctc	ctctcatggc	1860
gcctcatccc	ctctcttctc	ccccctccctc	tccctcctct	ctccctcctg	gccccatcct	1920
tctccacctt	cagctccact	atccccctct	ccctaccctg	tccttccctc	cttccgcctc	1980
cccccttccctc	ctcccgccca	ccgccccgca	cccgcccgtt	ccacccttcg	actttctcct	2040
gctgtggcct	aggctgagcc	gggagttacc	acttaactct	cactgggtct	ctcctgcacc	2100
ctatctctaa	acttctctcc	ttgggtgccc	cagctttcct	actcctgtct	ctcccgagct	2160
acctaggctt	ctctctctga	ctcyccgtct	ttctccagtt	atctacatct	gcagctcccc	2220
tgatgccttc	agaaccctca	tgtcctggc	cctggaagct	ggcttgtgtg	gggaggacta	2280
cgttttcttc	cacctggata	tctttgggca	aagcctgcaa	ggtggacagg	gccctgctcc	2340
ccgcaggccc	tgggagagag	gggatgggca	ggatgtcagt	gcccgccagg	cctttcaggt	2400
gagtacctag	gtttgaagcc	caggctgtct	cagcttgtgg	cacatcattt	ctgggcactg	2460
tgtccctcag	crtctgaaag	aattccagaa	aagaggtttt	tgtctgtttg	tttctttatg	2520
cactcctggt	aactcacaga	acagaaaaga	ggttggtgat	gctcactggg	aattaggcaa	2580
tgaagggcag	gggactgcc	aggggcgctt	cgccaccagc	aggctaaaaa	gataagaaaa	2640
tgggcttgag	gcgggaggag	gataaagtcc	cacagcctgg	acaggacttg	gagaaggcat	2700
ccatttgat	ccctgcttt	ggaatgggca	tcacttcatg	cagggcatag	ggtccagttt	2760
gaccttgagc	taagcagaga	cgcagctctg	ggaggtgggc	tcccaactgt	tggggcccca	2820
cagtactagg	gaatagtcag	ctcccaactc	tctgctctcc	actgaccctt	ttctcaggct	2880
gcaaaaatca	ttacatataa	agaccagat	aatcccagat	acttggaatt	cctgaagcag	2940
ttaaaacacc	tggcctatga	gcagttcaac	ttcaccatsg	aggatggcct	ggtaagaagg	3000
ggtcccggga	ccctccagcg	tggacctcca	gccccactc	catgaccctc	tgccagcctc	3060
catccttccc	tattccaggt	tctccccctc	cttccctccc	ttccattgt	tccatgtttc	3120
tcgtgatgat	ggaggaggac	actggcaagt	tcagcctctg	aaactcaggt	catcatcagt	3180
aatatggaga	cgatacatcc	tgcctgtct	acctagtagg	attcaggaag	tgatgctaata	3240
ccaaaggcat	cgtttaaata	gtaaaatctc	cctgtgatat	aggggtgtta	ttttctccca	3300
tcctcttcca	aaatcccagt	gcctcttggt	cccttcccca	cagctcccac	ctccatgccc	3360
ttcatatgcc	caccccagcc	gacctctggt	tgccctaca	ggtgaacacc	atcccagcat	3420
ccttcacga	cgggctcctg	ctctatatcc	aggcagtgac	ggagactctg	gcacatgggg	3480
gaactgttac	tgatggggag	aacatcactc	agcggtgtg	gaaccgaagc	tttcaagggtc	3540
agggcctgga	ggtggctgga	atgggctgcc	ttgggggatg	aatcccaggt	gcccagtgctc	3600

aagccatgag	aagcctattg	tcctgcagca	gttacctatg	cacaccagcc	ttttcctcca	3660
cagctttttt	caggcccatc	cctcagaagt	cctacaaagt	gtccaatctc	aatcatccct	3720
gctgggcact	gagttctttt	acctttcttt	ttcttttttc	tttttttttt	gagatggagt	3780
ctcgctctgt	ccccaaact	ggagtgtggt	ggtgcaatct	cggctcactt	caacctccgc	3840
ctcccagggt	caagcaattc	tcctgcctca	gcctcctgag	tagctgggat	tacagggtgcc	3900
ctccaccaac	acttggctaa	ttttttgtat	tttttttagt	agagacaggg	tttcaccacg	3960
ttggtcaggc	tggctctgaa	ctcctgacgt	caggatgatct	gcccgcctca	gcctcccaaa	4020
gtgctgggat	tacaagcatg	agccacagtg	cccggccggt	ttaccattta	ctatcattct	4080
gtatacatgt	atgtttggaa	ggcaaggcaa	aaaagattag	aggatgaaga	gatgaagtgg	4140
ggcacccttg	aacttctatt	ctctcaaaca	tagtcatctt	cccccatgtc	ctcagggtgtg	4200
acaggatacc	tgaaaattga	tagcagtggc	gatcgggaaa	cagacttctc	cctctgggat	4260
atggatcccg	agaatggtgc	cttcagggtg	agtttgtgca	cccagaagac	agtgccaat	4320
ccaaatgaca	tctcacctc	ctacwtcccc	cccacagccc	tgccagggca	cctgtttatc	4380
ctgtagccat	tccaccatgc	ctggacactt	acaagagccc	tggataaaa	agaccagct	4440
ccagtctggg	gaagccacca	gaatgatagg	gactcacag	catcacactt	ggggagcccc	4500
atgcctgagg	agggagcaca	agcctgcctt	cggggagctc	cgaaggagg	caggcaggac	4560
cgcctcccag	cagagacagg	gctgtgaaag	atgcacatta	cacagctctg	caagcgagca	4620
gggacaggaa	ggcgctgagg	ccaatggcca	caagggacag	gtcatccaga	gaaggcctcc	4680
tgaagacgg	gcacatggac	tgggcctgcg	aatgtaggct	aagggtgaaca	ttaccttctc	4740
ctgttttcta	ccaagaaaat	aagtagagaa	aatcaatgc	ttggttggtg	cttcaacca	4800
gattataaac	tccctgagtg	tagagatcgg	gttctaaatg	gagttttctt	tataaacccc	4860
ttgatagttt	tcagggtgtt	ccacttgagt	actatgtgtg	tggatgagg	tcctgtgtcc	4920
agttgcagt	gggacttgg	aagcagggtg	caaccagat	atatatgtag	gctctagaag	4980
cagagctggg	gtagggtggg	ggtgagactg	ctgcactcac	agcatgcctt	ccccgcaggc	5040
cctggcctag	ccaccactcc	tgtctccct	taggttgtac	tgaactaaa	tgggacttcc	5100
caagagctgg	tggctgtgtc	ggggcgcaaa	ctgaactggc	ccctggggta	ccctcctcct	5160
gacatcccca	aatgtggctt	tgacaacgaa	gaccagcat	gcaaccaagg	tgactgcccc	5220
ttgccttcca	ggcctccatc	ccagagatgc	tgcaccttc	ccctaagcac	agtcgagtag	5280
gtgctcctgy	cccatgctga	gggctttctg	gagaatgact	cctgcctttt	tcttcccttc	5340
atccatcatc	ccagttcact	gatggactat	tagaaagttc	ttcctcctgc	tgtctaacc	5400
aaatctctct	tgtctgaata	tggactctct	cctgcagatc	acctttccac	cctggagggtg	5460
ctggcttttg	tgggcagcct	ctccttgctc	ggcattctga	ttgtctcctt	cttcataaac	5520
aggtgagctg	tgatgtrggg	ggttgagtga	ggctggggga	cccggagaac	caagagcaga	5580
ggaggcggtg	gggaccaga	gggaagagg	caggggtgaa	ggggcagcag	gggaaaacca	5640
agggagatga	ggaagaaagg	aggcttaaaa	gccagaggag	aaagaaagag	aagggaatgg	5700
cagggcgagg	ggaggagaca	aggataggaa	tggccaagga	gagtcagaaa	gatccaagaa	5760
gcagagaagt	tgatgggtga	catcataggg	gcgtggactg	gttttccttg	ctactcttgc	5820
aggccagata	ggaagcaact	ttctgaacct	ttgcaatcat	gcccattgta	gctgaggagg	5880
gtgagccctg	gtgtgtgcca	ggtgccccac	ctagaatgga	gaagggagct	gaatgagcct	5940
tgttctgccc	gtccagtgga	ggctaaaatg	aagtacagga	ggagttaatg	atatacaaaa	6000
gcaaggaggg	aggggagaaa	aatcactgct	ggttgagcat	ataatgtgtg	ccaggcactt	6060
ccacgtacac	tatttctttc	tttctttttt	tttttttttt	tttttttttg	agaocggagtc	6120
tcgctctgtt	gccagactgg	agtgcagtgg	catgatctag	gctcactgca	acctccgcct	6180
cccagtttca	agcaattctc	ctgcctcagc	ctcccatgta	gctgggacta	caggcacatg	6240
ccaccacgct	cagctaattt	ttgtattttt	agtagagaca	gggtttcacc	atgttggcca	6300
ggatggtctc	gatctcttga	cctcatgato	caccacactt	ggcctcccaa	agtgtgggga	6360
ttacaggcat	gagccactgt	gcctggcctc	atgttcaacta	tttcttttca	ttcttataat	6420
agttaagaat	gaaatagata	ttgcggcctc	attcccaagt	aaggacattg	aggtgattcc	6480

```

cccaaggtcc ccagtaaggc agaatttccc ccagccatcc tgattctcag tccagaggat 6540
agaattcccc ctccatctct gagtgcattg tgtgggtccca cggctctgag gaggggctgc 6600
tgagcaccct gccctgggtc agcggctcag ccacaggctc agatgcagcc ttcgtatccc 6660
aggaagatgc agctggagaa ggaactggcc tcggagctgt ggcgggtgcg ctgggaggac 6720
gttgagccca gtagccttga gaggcacctg cggagtgcag gcagccggct gacctgagc 6780
ggggtaagaa cgctggtgtt tgtgttgggg ggcaataaag gagaggtggg tacaaggggc 6840
agtgcctgag ggataggtaa gggagcagga ttctagtccc agctctgctt tcaattgctr 6900
tgtgaccttg agcgactcat agtccctctc cgagactgtc tcagatgatg attacagcag 6960
cagagcctcc ctccacagggc tcttttaaa gtcagaggag atagtacctg tgaaaacact 7020
ttaaaaaaaa aaaaagtaaa tgaggaggaa attttatgat gtggaacata aagcagggtg 7080
ggccaggcac agtggctcac atctgcaatc ccagcacttt gggagaccga ggcaggagga 7140
ttgcttgtgc ctgggagttc aagaccagcc tgggcaacag agcaagacat cgtctctaca 7200
aagaatacaa agattagcag ggcattggtg cgcataacct tagtcccagc tactctggag 7260
gctgaggtga aaggatcatc tgagcccagg agtctgaggc ggcagtgacc taggatagca 7320
ccactgcact ccagcctgga tgacacaatg atactacatc tcaaaaaaaaa acccaacaac 7380
aaaaaggaag ggtgacacaa agataaggcw ggataaggca gggaaataaa gaccagagca 7440
caagcaatca ggatgcagac tgggcccacc ggctgaccat tctctctgct ctccctcctt 7500
tcagagaggc tccaattacg gctccctgct aaccacagag ggccagttcc aagtctttgc 7560
caagacagca tattataagg tgggcctggg gaaagatcac tgggccttgg gactggggca 7620
ggagtgtact ctgatggagg actggtgggg ggttctgagg gaaggagtaa gctggtgggg 7680
agcagcagat gggggccctg ggggtgggt atttggaaca agtgagggtc ctgagggcag 7740
ggatgggctg tcgggagcag ctggaattcc caggacatgg gacctgctc ttcacagtga 7800
cagtctccat tccatgccc gggcaacctc gtggctgtga aacgtgtgaa ccgtaaaccg 7860
attgagctga cacgaaaagt cctgtttgaa ctgaagcatg taatgtgggg agtgaggcag 7920
tggcatggag aaggggccct cggggaygca agggagactg gccaacagaa ctagtattgg 7980
agggacctca gggtagccca agaaaggggc agggactgga gccctggatg accttcatct 8040
tgtggtggag tgggggtatc ctaagtagga gaagagacca ctgagataac ctggaggaat 8100
cttgaggggc catatgtgat gtccctgggg gagagagggc ttaggatgcc agagggagta 8160
ggagcagatt ctggggaggg tgggctaaag gacatgggtg ggaatcacca gggaagatct 8220
tagtgatggt tgcagaaagt gaataaggag ttaagaagag tgagggtccc tgaagctagt 8280
gagcagcttg gtgaggagcg aggtctctgt caagctcctg atgctggtcc cacttgacga 8340
tgcgggatgt gcagaatgaa cacctgacca ggtttgtggg agcctgcacc gaccccccca 8400
atatctgcat cctcacagag tactgtcccc gtgggagcct gcaggtgagg gggacaaggg 8460
gtgtcaagaa acctgggttc tagccctggc tctgcccctg actggccata agaccccagg 8520
catgcctcgc cctctttctg acctttctgg ccccatctgt aaaaatggga gttggggaag 8580
ggcagtggca ctagagtcaa tccaaagttt tgtcctgttc taccagttca catcagtagg 8640
acctgcacc ctctccaac tcccaggggg atctgcaggg gattggtctt gactcttatt 8700
gccccagcag gacattctgg agaattgagag catcaccctg gactggatgt tccggtactc 8760
actaccaat gacatcgtca aggtatgccc ctaagcacct attggatgtg tagagcaggg 8820
gccaggcatg cttctcctgg ccaeggtgt aggtccact cctggccaat acctctgcc 8880
actcacattt ccagggcatt ctgtttctac acaatggggc tatctgttcc catgggaacc 8940
tcaagtcatc caactgcgtg gtagatgggc gctttgtgct caagatcacc gactatgggc 9000
tgagagactt cagggacctg gacccagagc aaggacacac cgtttatgcc agtgagcctt 9060
gactcttgaa cctaacacct gccccagca ccaccagta gggagactga tgcaaggcct 9120
ctgatgggct tgggcatgct tgtcctgact ccagcctcaa ttcattcacc catgaaaaag 9180
ggaaggccag acgaagtggg ttctaaggcc tcctctagct ctaacactct gtgatgcac 9240
cagatcagtt tcggccacac ccttgtttcc ccctcaccct ttagctttgg gctccctcac 9300
tcggtgacta ccracctctg acccacagaa aagctgtgga cggcccctga gctcctgcga 9360

```

atggcttcac	cccctgtgcg	gggctcccag	gctggtgacg	tatacagctt	tgggatcatc	9420
cttcaggaga	ttgccctgag	gagtgggggtc	ttccacgtgg	aagggtttgga	cctgagcccc	9480
aaaggtgaga	ggagcacacc	ttccttaaac	ccagccacag	tctcaacgaa	ccccagcccc	9540
agggagaggg	tcccctggca	gcaccaccac	accttccttc	tgtaatgggg	ttcagtcacc	9600
accctttgac	ccattgctrc	cagtgaccag	tcccccgccc	ccatgccttg	gtcttgact	9660
tcccctgcca	tctcwgetgg	ttgccccagt	ctctcactag	gcccttggcc	agccccaccc	9720
ctcagctcct	ctacccccca	atacagagat	catcgagcgg	gtgactcggg	gtgagcagcc	9780
ccccttcggg	ccctccctgg	ccctgcagag	tcacctggag	gagttggggc	tgctcatgca	9840
gcggtgctgg	gctgaggacc	cacaggagag	gccaccattc	cagcagatcc	gcctgacgtt	9900
gcgyaaattt	aacagggtccc	tggtgtttgt	catggatccc	ccaggccctt	cctccacagc	9960
caccattttac	ctaattgcttc	tggtctctggc	ttatcccagc	agtrgcagag	ggagaccact	10020
cacctcctcc	ctgtacatag	tcagctccag	ctcagcacag	cctcatgacc	ctcttcgcaa	10080
gtacagcatg	actcagctgt	ccccacagtc	ccctgccatt	catgcccctt	ccctccacca	10140
tcgacacccc	acacccttcc	tgccccactcg	ccttgctggc	ctctagactt	ctcagcagtg	10200
tgtaggatag	atgggcctcc	cgcctectgc	cctgtaggct	cttgccctc	cacgggagct	10260
cctgccccac	cccttgattt	cccttcccca	gcgtgcccac	caggcccagt	tctccagac	10320
acacccttct	gtggacatca	ctttgtccgc	aattgaccct	tgtcattctc	cacctccttt	10380
acctccttct	aactcactgg	gttcaacaaa	gatgaacaaa	atgtccatat	gtctgaagct	10440
tcatacttga	ccttggggtc	tcagaaaaga	attgaacttt	cttccttctg	ttttcccctg	10500
ctccccggta	tctgtctatg	ccctcaaccc	tgagcgtctc	tagagacctc	actgcagtct	10560
ggagggggaa	gtgcctaggg	gcgggcgctc	acgtaggctg	tgctgctcct	ctcttaccac	10620
ccccaccgcc	accctctgcc	cccagggaga	acagcagcaa	catcctggac	aacctgctgt	10680
cccgcatgga	gcagtacgcg	aacaatctgg	aggaactggt	ggaggagcgg	accaggcat	10740
acctggagga	gaagcgcaag	gctgaggccc	tgctctacca	gatcctgcct	cagtgagtgc	10800
ctgagtctgg	ggaccccccc	caacacaaaag	cccctgtccc	gacccccaac	tctgatcctg	10860
cacctgcctt	gaccccttag	ctcagtggct	gagcagctga	agcgtgggga	gacggtgcag	10920
gccgaagcct	ttgacagtgt	taccatctac	ttcagtga	ttgtgggttt	cacagcgtg	10980
tcggcggaga	gcacacccat	gcaggtaggc	cagggttcag	ccacagggtg	caggcaagct	11040
cagcatctgg	atcccaccag	ayctgccttc	tggttctgct	ttaccacact	gaccccagg	11100
ggggtccctt	acttctctgt	tctcttagct	tctcttccct	tccaggtggt	gacctgctc	11160
aatgacctgt	acacttgctt	tgatgctgtc	atagacaact	ttgatgtgta	caaggtagag	11220
gtgggagtgg	ggatgggaag	ggacagacag	acatggacaa	ggtcagaaaa	agatgagggg	11280
taggcagaat	gatgtggagt	cttaagagag	gagatcgggg	acacgggcag	agacagtgc	11340
acaggagac	ccgggaacag	gcagagaacc	catgtgggat	gggggatgag	caaagacaga	11400
tgagggtaca	gaatgacaga	cgctgcaccc	ggtgtgacgg	tgtggccggc	cgcacagttg	11460
cagccgtcaa	gtcctgcacc	ccctcgccac	tcccacaggt	ggagacaatt	ggcgatgcct	11520
acatggtggt	gtcagggtc	cctgtgcgga	acgggcggct	acacgcctgc	gaggtagccc	11580
gcatggccct	ggcactgctg	gatgctgtgc	gctccttccg	aatccgccac	cggccccagg	11640
agcagctgcg	cttgcgcat	ggcatccaca	caggtaaggc	cactgaagg	gcaggcgggc	11700
atccagaggc	caaggctttg	caagggaac	ttgtccctg	gccagcccc	tggccctttc	11760
atctctctct	ctctctctct	ctctctctct	ctctctctct	gtctctctct	ctctctctct	11820
ctctctctct	ctcacacaca	cacacacaca	cacacacaga	gctgggaact	cagatcctgc	11880
ctcctgcctg	tcttggtattg	tccacctacc	tcccttaaca	cccctccctc	cctcactcgc	11940
tgatgggtc	tgctccttcc	cttgctcctc	ccaggacctg	tgtgtgctgg	agtgggtggga	12000
ctgaagatgc	cccgttactg	tctctttggg	gatacagtca	acacagcctc	agaatggag	12060
tctaattggg	aaggtagagt	gccccctcct	agagggaatg	gggagggcag	ggtggctgag	12120
ggaaatgcca	tcctggggca	gcctgtgcct	gcacagcccg	tttcagctcc	tagccctttc	12180
gcctcccaag	ttccccctct	cataatatta	agagttcaac	ctgggctcat	caacttgact	12240

gtaaccagag	actcagggttc	ctgctgcccc	tcttgtcaaa	cgatgtaaaa	gtatttccgg	12300
gccagtgtg	gagagttccc	agcaggaatc	tgatttttaag	accctctgtg	ggccggggcgt	12360
ggtgactcac	acctgtgatc	ccagcacttt	gggaagctga	ggcagggcga	tcacctgagg	12420
tcgggggttt	cgagaccagc	ctgaccaaca	tgatgaaatc	ccgtctctac	taaaaataca	12480
aaaaactagc	cagggtgtgat	ggcaggctcc	tgtaatccca	gctacttggg	aggcttgagg	12540
cagaagaatt	gcttgaaccc	gggaggcaga	ggttgcgatg	agccaagatt	acaccacgca	12600
ccccagcttg	ggcaataaga	gttaaactct	gtctcaaaaa	aaaaaaaaaa	aaaaaaaaaa	12660
agggccctct	gctccacott	tgatgtggta	aagatggctt	cagagccagc	ataagtgagg	12720
ctgtgaatct	cagctccaca	gctggctgtg	tgctcagttt	ctatacctct	ctgagccatg	12780
gttttctca	tctgtaaaaa	gagggaaaaa	atctatctca	caggaattat	gtgagaaacc	12840
cattaaaaat	gtctaccaca	taattgtcat	ttaacttttc	caagccttag	cggattatct	12900
gtaaaatgat	gtctatctca	ggattgcaag	aagcctagca	caaaccctgg	taccagcag	12960
gcacctaata	aattcttact	cctaccogcc	ccttgctctt	gcctcctgtt	tatcttctat	13020
ccttctgctg	tattcgacac	aattcaatgc	agtaaacatt	tattgagtga	ctactgagt	13080
ccaggccctg	ggatagtaac	atggcccaga	tccagagtta	gctgagaaat	tcatgtggac	13140
cccatctaaa	ccttatgggtg	aaagaaaggc	tgcttgggag	ccagtcctgg	gagcccagag	13200
ggatctagtt	cggcaaatat	tccctgggca	ctatttgggg	gctgcagagt	cagcccttgt	13260
tgagggtcca	gtcctcaagg	agcacattcc	cagaaatgtt	cacattctgg	cgtggggtg	13320
ctgtaatccc	agcacttttg	gaggccgagg	tgggcagatc	acttgaggcc	aggagtggag	13380
actagcctgg	ccaacatggt	gacctoctgt	ctctactaaa	aatacaaaaa	attagctggg	13440
cgtggtggca	cgtgcccgtg	atcccagcta	ctcaggaggc	ttgagacatg	aaaatcactt	13500
gaaccagga	ggtggatggt	gcagtgagcc	gagactgcac	ccctgggcaa	cagagcgaga	13560
ctctgtctca	aaaaaaaaaa	agagagaaag	aaagaaaaga	aaagaaagaa	actgttaaac	13620
acaacaaggc	cactgtgatt	gatgcaaacc	ccagaagtag	ggacatgagt	tcagacagt	13680
gtcaaagaga	gggtgtggca	atattgggcc	ccactccatc	actgacctcc	tcagccactt	13740
gggcagatca	ccctgggcct	cagttcctcg	gccacaaaat	gagggtatag	catgaaatca	13800
tgaaagcaac	aattttacata	gtgcttcccta	ggtagcacat	tccgtttgaa	tactttatgg	13860
atgttaaatt	taatcctcac	aacaagggtt	tgagatgggt	actgacacta	tcagcatttt	13920
acagattagg	aaaatgaagc	agagagaatt	tattttacat	acctaagcaa	gtatccaagc	13980
tgaggttcat	actgaggcag	tgcaggatcc	aaagtgccag	ctcctaacca	ecatgctgtg	14040
tagagccggg	tgacactcca	gagagtgtct	tccaacagga	tgttccatag	tcatgaaaat	14100
gttctgtatt	ctgtgctgtc	caatacagta	gcctctaggc	acatatggct	acttatcact	14160
ggaaatgtga	cgggtgcaac	tgaggccctg	atTTTTTTTT	TTTTTTTgga	gacagagttt	14220
cgtctgtcg	cccagcctgg	atggagtgca	gtggtgcaat	ctcggctcac	tgcaacctcc	14280
gcctcccagg	ttcaagcgat	tctcctgcct	cagcctccca	agtagctgga	attacagggt	14340
agtgccacca	cacacagcta	atTTTTgtat	TTTTtagtaga	gacgggggtt	cgccatattg	14400
gccaggatgg	tctcgaaactc	ctggcctcaa	gtgatcctcc	tgccctcagcc	tcccaaagt	14460
ctgggattac	aggtgtgagc	cacagcacc	agcctgaatt	tttaactgta	tttagtttaa	14520
attaatttaa	gttgaaacag	gcacatgtga	ttagtggcta	ctgtattgga	ttacacagct	14580
ccagagttct	aaatgagagg	ctaattgtgt	cacgcactac	attcaggggg	tggggccct	14640
ctgagctaga	gggcttccctg	gccccaaaaga	gggagagagg	gtacctgtcc	acctgtccac	14700
ccccacagtc	cctgggtctct	tttgccctcta	ctttcctgct	ctcctctctc	acattgctca	14760
ccttcccttc	tcccctgtcc	taccagccc	tgaagatcca	cttgtcttct	gagaccaagg	14820
ctgtcctgga	ggagtgttgt	ggtttctgagc	tggagcttcg	aggggatgta	gaaatgaagg	14880
tagagcgaga	agcctctgcc	ctccccacct	tttgggggtcc	tagagggagt	taccttctc	14940
aagcagccga	tgccactccc	atccctaagg	ctctcatctg	actggggaaa	gggcatgtgc	15000
cactccccag	cccatacctct	tttttccctc	cagggcaaag	gcaaggttcg	gacctactgg	15060
ctccttgggg	agaggggggag	tagcaccocga	ggctgacctg	cctcctctcc	tatccctcca	15120

cacctcccct	accctgtgcc	agaagcaaca	gaggtgccag	gcctcagcct	cacccacagc	15180
agccccatcg	ccaaaggatg	gaagtaattt	gaatagctca	ggtgtgctga	ccccagtga	15240
gacaccagat	aggacctctg	agaggggact	ggcatggggg	gatctcagag	cttacaggct	15300
gagccaagcc	cacggccatg	cacagggaca	ctcacacagg	cacacgcacc	tgctctccac	15360
ctggactcag	gccgggctgg	gctgtggatt	cctgatcccc	tcccctcccc	atgctctcct	15420
ccctcagcct	tgctaccctg	tgacttactg	ggaggagaaa	gagtcacctg	aaggggaaca	15480
tgaaaagaga	ctaggtgaag	agagggcagg	ggagcccaca	tctggggctg	gccacaata	15540
cctgctcccc	cgaccccctc	cacccagcag	tagacacagt	gcacagggga	gaagaggggt	15600
ggcgcagaag	ggttgggggc	ctgtatgcct	tgcttctacc	atgagcagag	acaattaaaa	15660
tctttattcc	agtgacagtg	tctcttcttg	agggagagag	ggttgccaga	aaacagtcag	15720
ttctccactc	tctacttcaa	ataagactca	cttcttggtc	tacaagggtc	tagaaggaaa	15780
agtaaaaaaa	aaagactctc	gattcttaac				15810